

**UNIVERSITE MONTPELLIER 2  
SCIENCES ET TECHNIQUES DU LANGUEDOC**

**T H E S E**

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITE MONTPELLIER 2**

***Discipline : Biochimie et biologie moléculaire***

***Ecole Doctorale : Systèmes Intégrés en Biologie, Agronomie, Géosciences,  
Hydrosciences, Environnement***

présentée et soutenue publiquement

par

**Cyrille ZINI**

Le 21 Décembre 2010

**Structure d'un locus de résistance à la rouille chez une espèce  
hautement polyploïde, la canne à sucre ( $2n=ca\ 12x=ca\ 115$ )**

**JURY**

M. QUETIER Francis  
M. SOURDILLE Pierre  
M. CHARRIER André  
Mme GEFFROY Valérie  
Mme BERGES Hélène  
Mme D'HONT Angélique

Rapporteur  
Rapporteur  
Examineur  
Examineur  
Examineur  
Directeur de thèse



**UNIVERSITE MONTPELLIER 2  
SCIENCES ET TECHNIQUES DU LANGUEDOC**

**T H E S E**

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITE MONTPELLIER 2**

***Discipline : Biochimie et biologie moléculaire***

***Ecole Doctorale : Systèmes Intégrés en biologie, Agronomie, Géosciences,  
Hydrosciences, Environnement***

présentée et soutenue publiquement

par

**Cyrille ZINI**

Le 21 Décembre 2010

**Structure d'un locus de résistance à la rouille chez une espèce  
hautement polyploïde, la canne à sucre ( $2n=ca\ 12x=ca\ 115$ )**

**JURY**

M. QUETIER Francis  
N. SOURDILLE Pierre  
M. CHARRIER André  
Mme GEFFROY Valérie  
Mme BERGES Hélène  
Mme D'HONT Angélique

Rapporteur  
Rapporteur  
Examineur  
Examineur  
Examineur  
Directeur de thèse



## Remerciements

Je tiens à remercier en tout premier lieu Angélique D'Hont qui a dirigée cette thèse et sans qui tout cela n'aurait été pas possible. Je la remercie pour sa patience et ses conseils, pour m'avoir fait découvrir le travail sur une plante à génome complexe.

Je remercie Mr Francis Quétier et Mr Pierre Sourdille pour avoir jugé ce travail en tant que rapporteur. Je remercie aussi Mme Valérie Geffroy, Mme Hélène Berges et Mr André Charrier pour leur participation au jury de thèse.

Je remercie la région Languedoc-Roussillon et le CIRAD d'avoir financés cette thèse.

Je remercie Carine Charron pour son aide apportée dans cette thèse et surtout dans la rédaction de cette thèse. Un très grand merci à toi.

Je remercie Olivier Garsmeur de m'avoir formé à toutes sortes de technique durant cette thèse.

Je remercie Catherine Hervouet qui m'a aussi accompagné tout le long de cette thèse en m'aidant dans les manip.

Je remercie l'ensemble de l'équipe Structure et Evolution des génomes : Nabila, Francky, Maguy, Céline, Valérie et ceux qui ont été de passage : Audrey, Marlène, Solen.

Il me reste beaucoup de personnes à remercier alors je dis un grand merci à l'ensemble du personnel de l'UMR Développement et Amélioration des Plantes. Merci à vous tous et à tous les gens que j'ai pu rencontrer lors de cette thèse. Une mention spéciale tout de même à Julien dit le jeune.

Cela va de soi, je remercie évidemment ma famille pour son irremplaçable et inconditionnel soutien. Ils ont été présents pour écarter les doutes, soigner les blessures et partager les joies. Cette thèse est un peu la leur, aussi



Le plus fort de mes remerciements est pour mon épouse Angélique. Merci de m'avoir tenu la main jusqu'aux dernières lignes de ce mémoire. En plus de me supporter dans des conditions «normales», ce qui n'est déjà pas une mince affaire je le conçois, elle a du subir mon humeur de barbu en fin de thèse et se mettre en retrait pour s'occuper de notre fils. Et pourtant, quelle présence indispensable. Merci d'être là tous les jours.





# Sommaire

<b>Chapitre 1 : Introduction</b>	<b>1</b>
<b>A. La canne à sucre</b>	<b>2</b>
A.1. Importance économique	2
A.1.1. La canne à « sucre »	2
A.1.2. La canne à « éthanol »	3
A.2. Biologie et physiologie de la canne à sucre	4
A.3. Le complexe <i>Saccharum</i>	5
A.3.1. Les espèces sauvages	5
A.3.2. Les espèces cultivées	6
A.4. Evolution et domestication de la canne à sucre	7
A.5. Structure des génomes des cultivars modernes	9
A.6. Cartes génétiques de canne à sucre disponibles	10
A.7. Relations synténiques entre génomes de Poacées	11
<b>B. Les champignons de la rouille</b>	<b>12</b>
B.1. Caractéristiques générales et taxonomie	12
B.2. La rouille brune chez la canne à sucre : <i>Puccinia melanocephala</i>	13
B.2.1. Historique et distribution	13
B.2.2. Races de <i>P. melanocephala</i>	14
B.2.3. Symptômes et transmission de la rouille brune	15
<b>C. Les interactions plantes / microorganismes phytopathogènes</b>	<b>16</b>
C.1. Mécanismes de résistance	16
C.2. Les gènes <i>R</i> de résistance	17
C.3. Gènes de résistance à la rouille chez les Poacées	18
<b>D. Clonage positionnel chez les plantes à grand génome</b>	<b>20</b>
D.1. Contraintes du clonage positionnel chez les plantes à grand génome	20
D.1.1. La taille du génome	22
D.1.2. La redondance génétique	22
D.2. Stratégies utilisées/déployées pour contourner ces contraintes	23
D.2.1. Utilisation d'un donneur diploïde ou d'un parent proche diploïde à plus petit génome	24
D.2.2. Synténie avec les plantes modèles	24
D.2.3. Analyse de Ségrégation en mélange	25
<b>E. Etape et stratégies mis en place pour le clonage positionnel de <i>Bru1</i></b>	<b>26</b>
<b>F. Organisation du locus de <i>Bru1</i></b>	<b>29</b>
<b>G. Objectifs de la thèse</b>	<b>31</b>
<b>Chapitre 2. Cartographie physique de l'haplotype cible</b>	<b>33</b>
<b>A. Stratégies</b>	<b>34</b>
A.1. Stratégies générales	34
A.2. Ressources BAC utilisées et développées	35
A.3. Principe d'identification des clones BAC correspondant à l'haplotype cible	36
<b>B. Cartographie physique pour compléter l'espace de gauche</b>	<b>37</b>
B.1. Long-range PCR	38
B.2. Recherche d'une région orthologue chez le sorgho correspondant à l'insertion sur l'haplotype cible	38



B.3.	Criblage des banques BAC HindIII	39
B.3.1	Côté droit de l'espace de gauche	39
B.3.2	Côté gauche de l'espace de gauche	39
B.4.	Criblage de la banque BamHI	40
B.4.1.	Criblage de la première banque BamHI constituée de 40 000 clones	41
B.4.2.	Criblage de la banque BamHI étendue à 119 040 clones	43
B.4.3.	Marche sur chromosome à partir des trois banques BAC	45
B.5.	Conclusion sur la carte physique de l'espace de gauche	46
<b>C.</b>	<b>Cartographie physique pour compléter l'espace de droite</b>	<b>47</b>
C.1.	Eliminer l'espace de droite par réduction de la région cible	47
C.2.	Criblage des banques HindIII	48
C.3.	Criblage de la banque BamHI	49
C.4.	Marche sur chromosome à partir des trois banques BAC	50
C.5.	Conclusion sur la carte physique de l'espace de droite	51
<b>D.</b>	<b>Séquençage et annotation des nouveaux BAC de l'haplotype cible</b>	<b>51</b>
<b>E.</b>	<b>Conclusion</b>	<b>54</b>
<b>Chapitre 3 : Analyse des gènes candidats présents sur la séquence partielle de l'haplotype cible</b>		<b>56</b>
<b>A.</b>	<b>Méthodes</b>	<b>57</b>
A.1.	Méthodes d'analyses de la structure des gènes	57
A.2.	Extraction des ARN	58
A.3.	Obtention des ADNc	59
A.4.	Amplification par PCR	59
<b>B.</b>	<b>Endoglucanase de la région de Bru1 (gène 12)</b>	<b>59</b>
B.1.	Implication des endoglucanases dans les mécanismes de résistance	59
B.2.	Analyse structurale de l'endoglucanase de la région de <i>Bru1</i>	61
<b>C.</b>	<b>Kinases présentes sur l'haplotype cible (gènes 6, 10, 15, 16)</b>	<b>61</b>
C.1.	Implication des kinases dans les mécanismes de résistance	61
C.2.	Analyse structurale des kinases présentes sur l'haplotype cible	63
C.2.1	S/T kinase codée par le gène 6 (BAC CIR9O20)	63
C.2.2	S/T kinase codée par le gène 10 (BAC CIR9O20)	64
C.2.3	S/T kinase codée par le gène 15 (contigs 22M6_CIR12E03)	64
C.2.4	S/T kinase codée par le gène 16 (contigs 22M6_CIR12E03)	66
C.3.	Analyse fonctionnelle de la kinase candidate codée par le gène 15	67
C.3.1.	Test d'expression du gène 15	68
C.3.2.	Conclusion	69
<b>Chapitre 4 : Origine de l'insertion contenant le gène Bru1</b>		<b>70</b>
<b>A.</b>	<b>Matériels et Méthodes</b>	<b>72</b>
A.1.	Matériel végétal.	72
A.2.	Marqueurs PCR	72
A.2.1.	Marqueurs PCR dérivés de la carte physique de <i>Bru1</i>	72
A.2.2.	Marqueurs ISBP	73
A.2.3.	Marqueurs microsatellites	74
<b>B.</b>	<b>Distribution des marqueurs dans la collection</b>	<b>74</b>
B.1.	Distribution des marqueurs dans la généalogie du cultivar moderne R570	74



B.2. Distribution des marqueurs dans la collection de clones des espèces ancestrales.	76
C. Conclusion sur l'origine de l'insertion	77
<b>Discussion générale et perspectives</b>	<b>79</b>
A. Cartographie physique de l'haplotype porteur du gène <i>Bru1</i>	80
B. Gènes candidats	82
C. Origine de l'insertion	83
<b>Conclusion générale</b>	<b>89</b>
<b>Références Bibliographiques</b>	<b>91</b>
<b>Annexe 1.</b>	<b>104</b>
<b>Annexe 2.</b>	<b>112</b>



## **Sommaires des Figures et Tableaux.**

### **Chapitre 1 : Introduction**

**Tableau 1.1.** Principaux pays producteurs de canne à sucre en 2009

**Figure 1.1.** Taxonomie du complexe *Saccharum*

**Figure 1.2.** Aires d'origine et de dispersion des espèces du genre *Saccharum*

**Figure 1.3.** Scénario d'évolution et de domestication de la canne à sucre

**Figure 1.4.** Hybridation génomique *in situ* (GISH) d'un clone de *S. barberi* ( $2n = 82$ )

**Figure 1.5.** Transmission des chromosomes au cours de la nobilisation pour l'obtention du clone POJ 2878

**Figure 1.6.** Structure du génome du cultivar R570 ( $2n = \text{ca } 115$ ).

**Tableau 1.2.** Cartes génétiques disponibles chez la canne à sucre

**Figure 1.7.** Synténie entre les principaux génomes de céréales

**Figure 1.8.** Relations phylogénétiques entre les principales espèces de graminées

**Figure 1.9.** Conservation des gènes entre des régions orthologues de la canne à sucre, du sorgho, du riz et deux régions hom(é)ologues du maïs

**Figure 1.10.** Cycle de vie d'une rouille de blé : *Puccinia graminis*

**Figure 1.11.** Arbre phylogénétique d'un échantillon de champignons appartenant au genre *Puccinia*

**Figure 1.12.** Exemples de symptômes causés par l'agent de la rouille brune *P. melanocephala* chez la canne à sucre

**Figure 1.13.** Modèles biochimiques de la relation gène-pour-gène

**Figure 1.14.** Les grandes classes de protéines codées par les gènes de résistance *R* chez les plantes

**Figure 1.15.** Schématisation des étapes du clonage positionnel d'un gène d'intérêt

**Tableau 1.3.** Exemples de banques BAC pour quelques espèces de Poacées

**Tableau 1.4.** Ratio de ségrégation d'un marqueur simplex, duplex, triplex dans une descendance issue d'une autofécondation, pour différents niveaux de polyploïdie ( $h$ ) et d'appariements des chromosomes

**Figure 1.16.** Relations synténiques entre le groupe de liaison VII du cultivar R570 et les régions orthologues chez le sorgho, le maïs et le riz

**Figure 1.17.** Exploitation de la synténie avec le sorgho et le riz pour saturer la carte génétique de la région de *Bru1*





**Figure 1.18.** Carte physique de la région de *Bru1*

**Figure 1.19.** Comparaison de l'organisation physique entre sept clones BAC hom(é)ologues de canne à sucre avec les régions orthologues du sorgho et du riz de la région de *Bru1*

## **Chapitre 2 : Cartographie Physique de l'haplotype cible**

**Figure 2.1.** Carte physique de la région de *Bru1*

**Figure 2.2.** Représentation de l'haplotype cible dans les 3 banques BAC

**Figure 2.3.** Organisation physique des 7 clones BAC hom(é)ologues au niveau de la région de *Bru1* et des segments orthologues chez le sorgho et le riz

**Figure 2.4.** Visualisation du résultat du BLASTN avec le programme Circos obtenu d'après les séquences nucléotidiques des gènes présents dans l'insertion de l'haplotype cible et la séquence du génome total de sorgho disponible dans les bases de données

**Figure 2.5.** Carte physique et génétique de l'espace de gauche avec l'annotation des BACs déjà identifiés avant la thèse

**Figure 2.6.** Carte physique et génétique de l'espace de droite avec l'annotation des BACs déjà identifiés avant la thèse

**Figure 2.7.** Annotation des nouveaux BACs identifiés appartenant à l'haplotype cible

**Figure 2.8.** Annotation finale des BAC hom(é)ologues de la région de *Bru1* et des BAC des régions orthologues du sorgho et du riz

## **Chapitre 3 : Analyse des gènes candidats présents sur la séquence partielle de l'haplotype cible**

**Figure 3.1.** Localisation des gènes candidats sur les séquences des BACs de la région de *Bru1*

**Figure 3.2.** Alignement des séquences en acides aminés des protéines S/T kinases de canne à sucre, de riz et d'orge

**Figure 3.3.** Alignement des séquences en acides aminés de l'endoglucanase présente sur l'haplotype cible de la région de *Bru1* de canne à sucre et des  $\beta$ -1,4- endoglucanases et des  $\beta$ -1,3- endoglucanases présentes chez le sorgho, le riz, le peuplier et l'arabette

**Figure 3.4.** Arbre phylogénique obtenu à partir de l'alignement des séquences en acides aminés des endoglucanases de canne à sucre, sorgho, riz et peuplier



**Figure 3.5.** Séquences présentes sur les BACs de la région cible de *Bru1* chez la canne à sucre présentant des homologies de séquences avec des sérine/thréonine kinases

**Figure 3.6.** Arbre phylogénétique obtenu à partir de l'alignement des séquences en acides aminés des protéines orthologues à la protéine codée par le gène 15

**Figure 3.7.** Structure du gène 15 codant la kinase candidate et localisation des couples d'amorces définis pour les tests d'expression

**Figure 3.8.** Tests d'expression des ADNc du gène 15

## **Chapitre 4 : Origine de l'insertion contenant le gène *Bru1***

**Figure 4.1.** Origine des marqueurs moléculaires sur le groupe d'homologie VII du cultivar R570

**Figure 4.2.** Localisation des marqueurs PCR et ISBP sur la séquence de l'haplotype cible de la région de *Bru1*

**Figure 4.3.** Exemple de profils PCR obtenus avec le marqueur ISBP Sc1080

**Figure 4.4.** Distribution des marqueurs 9O20D11/RsaI\* et R12H16\* dans la généalogie du cultivar R570

**Figure 4.5.** Profil des marqueurs PCR et ISBP sur la collection de clones de canne à sucre

**Tableau 4.1.** Profils PCR et nombre de profils microsatellites de 26 clones *Saccharum*

**Figure 4.6.** Phylogénie de 13 clones de *S. barberi* et *S. sinense* basée sur 203 marqueurs RFLP révélés par une combinaison de 10 sondes / enzyme

## **Chapitre 5 : Discussion générale et perspectives**

**Figure 5.1** Représentation schématique de la différenciation des haplotypes de canne à sucre basée sur la comparaison des séquences des allèles hom(é)ologues



## **Liste des abréviations par ordre d'apparition**

RFLP	Restriction Fragment Length Polymorphism
GISH	Genomic In Situ Hybridization
APPCR	Arbitrary Primed Polymerase Chain Reaction
AFLP	Amplified Fragment Length Polymorphic
SSR	Simple Sequence Repeat
RAF	Random Amplified Fingerprinting
cM	centiMorgan
kb	Kilo bases
Mb	Méga bases
Gb	Giga bases
GC	Groupe de Coségrégation
PCR	Polymerase Chain Reaction
Gène <i>R</i>	gène de Résistance
Gène <i>Avr</i>	gène d'Avirulence
LRR	Leucine Rich Repeat
NBS	Nucleotide Binding Site
CC	<u>C</u> oiled <u>C</u> oiled
LZ	<u>L</u> eucine <u>Z</u> ipper
TIR	<u>T</u> oll and <u>I</u> nterleukin-1 <u>R</u> eceptors
TM	domaine TransMembranaire
S/T kinase	Sérine / Thréonine kinase
BAC	Bacterial Artificial Chromosome
YAC	Yeast Artificial Chromosome
TE	Elément Transposable
ADN	Acide DésoxyriboNucléique
ADNc	Acide Désoxyribonucléique complémentaire
ADNg	Acide Désoxyribonucléique génomique
BSA	Bulk Segregant Analysis
EST	Expressed Sequence Tag
LTR	Long Terminal Repeat
BES	Séquences d'Extrémités de BAC
CDS	Coding Sequence Domain
CAPS	Cleaved Amplified Polymorphic Sequence
HP	Hypothetical Protein
BLAST	Basic Local Alignement Search Tool
ARN	Acide RiboNucléique
ARNm	Acide RiboNucléique Messenger
GOST	GreenPhyl Orthologs Search Tool
ISBP	Insertion Site-Based Polymorphism
BR	Bulk Résistant
BS	Bulk Sensible



---

# **Chapitre 1 : Introduction**

---

**Tableau 1.1. Principaux pays producteurs de canne à sucre en 2009**(statistiques FAO, <http://faostat.fao.org/>)

<b>Pays</b>	<b>Production (millions de tonnes)</b>	<b>Surface cultivée (hectares)</b>
Brésil	690	8 598 440
Inde	285	4 400 000
Chine	114	1 630 520
Thaïlande	67	932 465
Mexique	51	669 231
Pakistan	50	1 029 000
Colombie	38	379 505
Australie	31	391 291
Argentine	30	355 000
USA	27	353 659
Indonésie	26	420 000
Philippines	23	404 000
La Réunion	1,8	24 528
Guadeloupe	0,7	10 500
Monde	1 682	23 727 751



## Chapitre 1 : Introduction

### **A. La canne à sucre**

#### **A.1. Importance économique**

##### **A.1.1. La canne à « sucre »**

La canne à sucre est à la base de l'industrie sucrière mondiale. C'est autour de cette culture qu'est née l'industrie d'extraction du sucre, en Chine ou en Inde, entre 1500 et 1000 ans avant notre ère. En 2009, la production mondiale de canne à sucre a dépassé les 1,6 milliards de tonnes, sur une surface cultivée totale de près de 24 millions d'hectares, représentant les trois quarts de la production totale en saccharose. Le quart restant provient majoritairement de la betterave sucrière. Cent quatorze pays produisent du sucre et parmi eux 76 ne cultivent que de la canne à sucre. Parmi ces pays, 12 totalisent plus de 80 % de la production mondiale (Tableau 1.1). Ces pays se situent entre le 35° parallèle de latitude Nord et le 30° parallèle de latitude Sud. Le Brésil, l'Inde et la Chine sont les plus gros producteurs de canne à sucre (source FAO 2009, <http://faostat.fao.org/>). A plus petite échelle, La Réunion est la première région productrice de sucre de canne en Europe. En Guadeloupe, la filière canne à sucre représente le secteur dominant de l'industrie agroalimentaire de l'île.

Le saccharose n'est pas le seul dérivé obtenu après transformation de la canne à sucre. Les autres principaux dérivés sont :

- la bagasse : résidu fibreux obtenu après extraction du jus, composé principalement par la cellulose de la plante. Elle sert de combustible dans les sucreries et rentre dans l'alimentation animale.
- la mélasse : sous-produit final de l'extraction du sucre. C'est un sirop très épais et très visqueux constituant un résidu du raffinage du sucre. Elle est utilisée dans la production du rhum industriel et parfois dans l'alimentation animale.
- l'écume : résidu de filtration. Elle est utilisée comme engrais pour la culture de la canne à sucre.



### A.1.2. La canne à « éthanol »

La canne à sucre est également une espèce majeure pour la production de biocarburant à partir du sucre, de la mélasse et de la bagasse. La moitié de la production de sucre au Brésil est destinée à la fabrication de bioéthanol, avec un rendement de 7 500 litres par hectare par an (Waclawovsky *et al.*, 2010). Le bioéthanol est aujourd'hui produit à partir de plantes sucrières (canne à sucre, betterave) et de plantes amylacées (blé, maïs). Cette filière « alcool » passe par une étape de fermentation par des levures du genre *Saccharomyces* qui transforment les sucres en éthanol, et une étape plus ou moins poussée de distillation qui sépare l'alcool de l'eau.

Une étude récente menée par De Vries *et al.*, (2010) a comparé l'utilisation de 9 plantes pour la production de biocarburants : 6 plantes pour la production de bioéthanol (maïs, blé, betterave à sucre, manioc, sorgho à sucre et canne à sucre) et 3 plantes pour la production de biodiesel (palmier à huile, soja et colza). Différents facteurs/paramètres ont été testés/mesurés comme le rendement énergétique, l'utilisation d'eau, l'émission de gaz à effet de serre... Le point fort de la canne à sucre par rapport aux autres espèces testées est son rendement énergétique élevé, avec un ratio de 1:8 (1 unité énergétique nécessaire à la culture de la canne pour 8 unités d'énergie produite). Ainsi, la canne à « énergie » est en croissance spectaculaire sur la planète.

En revanche, pour conserver la couverture mondiale des besoins alimentaires en sucre, la production à grande échelle de bioéthanol impliquerait d'augmenter les surfaces agricoles et donc, comme au Brésil, de procéder à une déforestation massive. C'est pourquoi des recherches ont été entreprises sur les carburants de deuxième génération, dont la production est beaucoup plus rentable sur le plan énergétique et qui préserve davantage l'environnement. Dans le cas de la canne à sucre, il s'agirait alors de valoriser la plante entière par fermentation enzymatique de la cellulose. Il n'y a pas encore de carburant de seconde génération dérivé de la canne à sucre commercialisée.

Des programmes de sélections ont été entrepris chez différentes graminées, dont la canne à sucre, pour augmenter le rendement de la biomasse ligno-cellulosique car la qualité et la quantité des lignines dans les parois, et les liaisons entre lignines et glucides pariétaux,



limitent les rendements de transformation en carburant par les micro-organismes fermenteurs (Lam *et al.*, 2009; De Vries *et al.*, 2010; Waclawovsky *et al.*, 2010). L'objectif de ces travaux est de s'approcher de la description d'un idéotype de paroi végétale facilement dégradable en bioéthanol, en s'appuyant sur des démarches de biochimie, de génétique et de génomique.

## **A.2. Biologie et physiologie de la canne à sucre**

La canne à sucre est une plante en C4, considérée comme une des plus productives en biomasse. C'est une graminée herbacée pérenne composée de talles ou tiges, dressées et regroupées en touffes. La tige de la canne à sucre (organe de stockage du saccharose) compte 20 à 30 entre-nœuds à maturité et mesure entre 2 à 5 mètres pour un diamètre de 2 à 5 cm. Chaque nœud correspond au point d'insertion d'une feuille et comporte un bourgeon axillaire. Sa propagation se fait essentiellement par bouturage de tige à partir du développement des bourgeons axillaires : les tiges coupées en morceaux et contenant au minimum un bourgeon sont mises en terre afin d'obtenir une nouvelle plante. A maturité, la tige développe à son sommet une inflorescence terminale pouvant atteindre 1 m de long. Les ramifications portent des paires de petits épis (épillets) qui donnent des graines. L'ensemble des graines d'une panicule est appelé le « fuzz ». La reproduction sexuée est réservée à la création variétale.

Les conditions environnementales idéales pour le bon développement de la plante reposent essentiellement sur une bonne irrigation et une température moyenne de 30 à 34°C tout au long de la période de croissance avec une période plus sèche et plus froide avant la récolte. L'accumulation du saccharose dans la tige a lieu à la fin de la période de végétation, après la floraison. Elle est déclenchée par une action combinée du froid, même relatif, et d'une diminution de l'alimentation hydrique. En cas d'absence de saison sèche très marquée, la richesse en sucre des tiges est faible. Généralement, la canne est récoltée avant que la plante ne fleurisse pour ne pas baisser la teneur en sucre des tiges et peut se faire 10 à 24 mois après la mise en terre de la bouture. La canne repousse après chaque coupe mais le rendement décroît après chaque cycle de récolte.

Famille	Sous-Famille	Tribu	Groupe de sous-tribu	Sous-Tribu	Genre
Poaceae	Panicoideae	Andropogoneae	Dimeriinae	Saccharastrae	Imperata
			Saccharinae		Eriochrysis
	Paniceae		Germainiinae	Eulaliastre	Eccoilopus
			Arthraxoninae		Spodiopogon
			Andropogoninae		Miscanthidium
			Anthistirinae		<b>Erianthus</b>
			Ischaeminae		<b>Miscanthus</b>
			Rottboeliinae		<b>Sclerostachya</b>
			Tripsacinae		<b>Narenga</b>
			Coicinae		<b>Saccharum</b>
			Chinachninae		
	Chlorideae				
	Bambusoideae				
	Pooidae				

**Figure 2.1. Taxonomie du complexe *Saccharum*.**

Les genres indiqués en gras font partie du complexe *Saccharum* (Daniels & Roach, 1987).

### A.3. Le complexe *Saccharum*

La canne à sucre est une monocotylédone de la famille des Poacées, comme le riz, et de la tribu des Andropogonées au même titre que le maïs et le sorgho (Figure 1.1). L'ensemble des données acquises ces vingt dernières années grâce aux outils moléculaires montre que la canne à sucre appartient au genre monophylétique *Saccharum* constitué de cinq espèces de base, parmi lesquelles on distingue les espèces sauvages et les espèces cultivées (Grivet *et al.*, 2006). Toutes les espèces du genre *Saccharum* ont un génome hautement polyploïde.

#### A.3.1. Les espèces sauvages

*Saccharum spontaneum* est une espèce rustique que l'on retrouve dans des régions très contrastées (températures, humidités ou types de sols). Cette espèce est présente sur une étendue géographique s'étalant du Japon jusqu'à l'Afrique de l'Est en passant par l'Inde et le Moyen-Orient (Brandes *et al.*, 1939). *S. spontaneum* présente des tiges de petits diamètres qui tallent abondamment. Les tiges contiennent peu ou pas de sucre mais sont très riches en fibres. La reproduction se fait par voie sexuée ou par voie végétative. Il existe des espèces annuelles, semi-pérennes ou pérennes. *S. spontaneum* présente des rhizomes.

Le nombre de chromosomes varie de  $2n = 40$  à  $2n = 128$  (Sreenivasan *et al.*, 1987). D'Hont *et al.*, (1998) ont montré, par hybridation *in-situ* des ADN ribosomaux, que le nombre de base est  $x = 8$ , et que le niveau de polyploïdie varie entre 8 et 12 selon les cytotypes étudiés.

*Saccharum robustum* est une espèce endémique de Papouasie-Nouvelle-Guinée. On différencie *S. robustum* de *S. spontaneum* par l'absence de rhizomes, une inflorescence plus grande, des épillets plus petits et une biomasse plus importante due à des tiges généralement plus longues, plus grosses et plus ligneuses (Stevenson, 1965). Les tiges ne contiennent quasiment pas de sucre.

On trouve deux cytotypes euploïdes :  $2n = 60$  ou  $2n = 80$  et des aneuploïdes variants de  $2n = 63$  à 205. Ces clones aneuploïdes pourraient correspondre à des hybrides





naturels entre clones de *S. robustum* ou entre *S. robustum* et d'autres espèces (Price, 1957; Price, 1965). Le nombre chromosomique de base est de  $x = 10$  (D'Hont *et al.*, 1998).

### A.3.2. Les espèces cultivées

*Saccharum officinarum* est considérée comme l'espèce sucrière, que l'on appelle encore espèce « noble ». Cette espèce présente de grosses tiges très riches en saccharose, et possédant une faible teneur en fibre et en amidon (Bremer, 1961; Roach, 1986). Bien que considérée initialement comme très homogène, *S. officinarum* présente des caractéristiques agronomiques qui sont en réalité assez variables suivant les clones (Daniels & Roach, 1987), comme l'étaye par ailleurs une récente étude de diversité génétique menée à l'aide de marqueurs moléculaires (Jannoo *et al.*, 1999). Cette espèce correspond aux formes domestiquées, dont les plus intéressantes ont été propagées en culture pendant des siècles jusqu'à l'intervention de l'hybridation interspécifique, notamment pour contrer les problèmes de maladies.

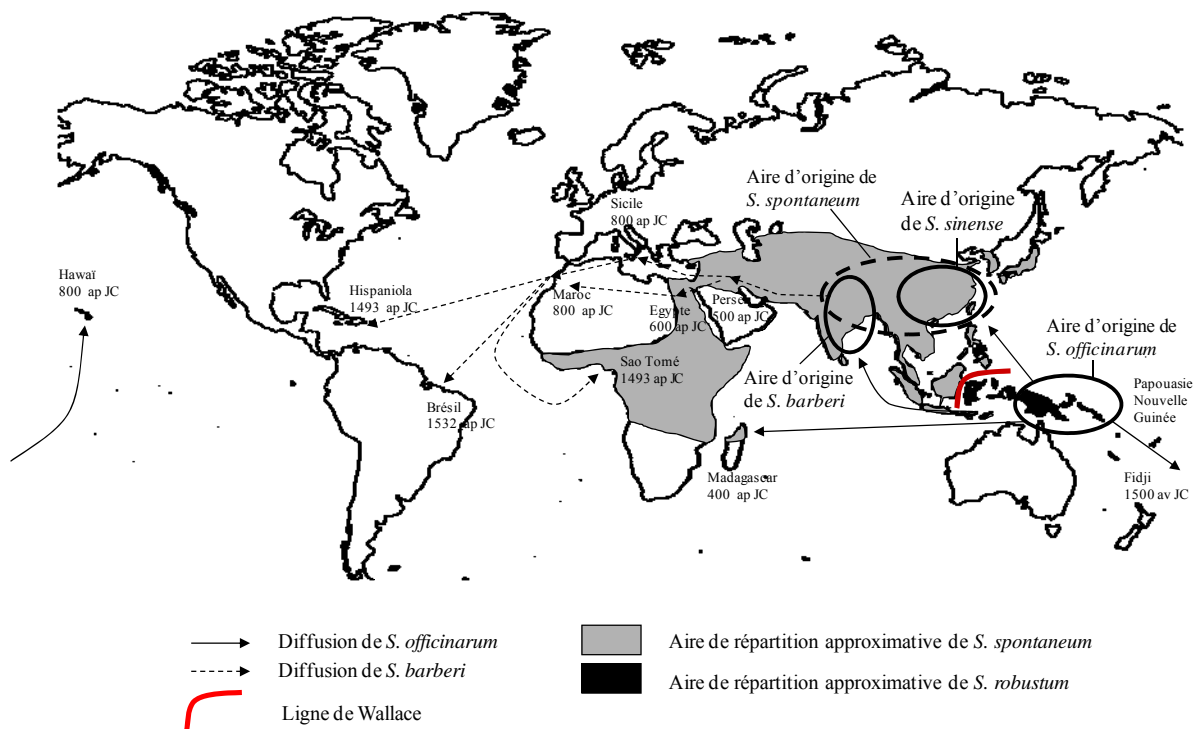
*S. officinarum* est une espèce euploïde à  $2n = 80$  chromosomes (Price, 1968). D'Hont *et al.*, (1998) ont montré que cette espèce possède un nombre chromosomique de base de  $x = 10$ .

*Saccharum barberi* et *Saccharum sinense* sont deux espèces sucrées qui ont été cultivées respectivement en Inde et en Chine. Elles se distinguent de *S. officinarum* par leurs caractères floraux, leur haute teneur en fibre et leur meilleure tolérance aux stress environnementaux.

Le nombre chromosomique de ces deux espèces varie de  $2n = 82$  à  $2n = 124$ .

*Saccharum edule* est cultivée de la Nouvelle-Guinée jusqu'aux îles Fidji pour ses inflorescences avortées comestibles. Ses tiges sont longues mais ne possèdent pas de sucre.

Le nombre chromosomique varie de  $2n = 60$  à  $2n = 122$ .

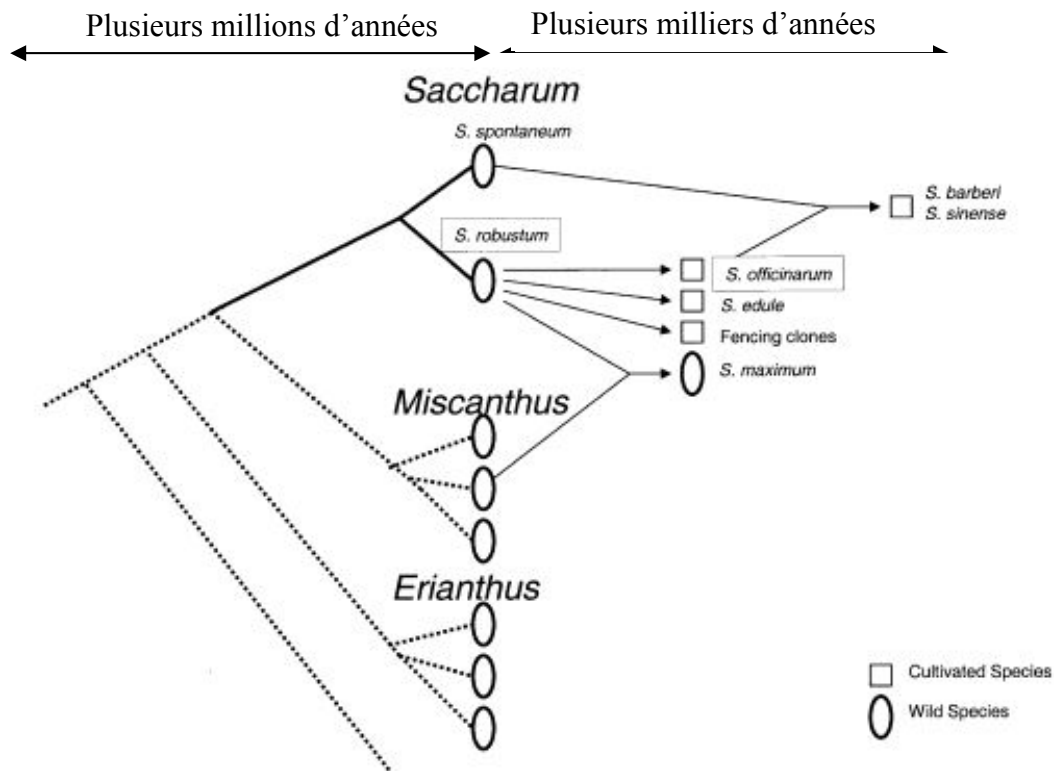


**Figure 1.2. Aires d'origine et de dispersion des espèces du genre *Saccharum* (d'après Blume, 1985; Daniels and Roach, 1987; Meyers, 1989; Panje et Babu, 1960).**

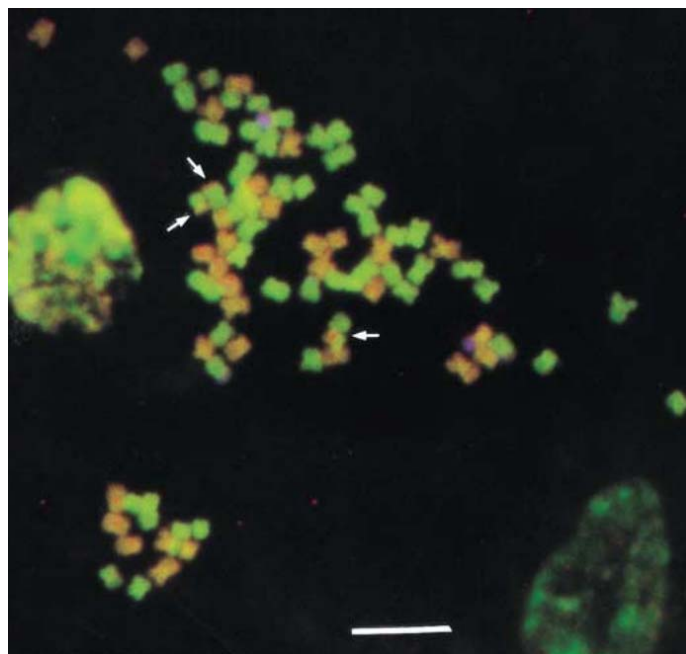
#### A.4. Evolution et domestication de la canne à sucre

La domestication de la canne à sucre aurait eu lieu en Papouasie-Nouvelle-Guinée environ 2 500 ans avant notre ère (Figure 1.2). La ligne de Wallace est une frontière théorique qui sépare l'empire biogéographique indomalais (Asie du Sud-Est, Java, Sumatra) de l'empire biogéographique australien (Papouasie-Nouvelle-Guinée et les îles voisines), empires restés séparés par les eaux pendant plus de trois millions d'années. Les espèces sauvages de *Saccharum* se seraient différenciées de part et d'autres de cette ligne : *S. spontaneum* du côté indomalais et *S. robustum* du côté australien. Ces deux espèces ont été clairement différenciées par marquage moléculaire en utilisant des sondes RFLP (Restriction Fragment Length Polymorphism) (Lu *et al.*, 1994b). La domestication de *S. officinarum* s'est probablement déroulée en Papouasie-Nouvelle-Guinée à partir de l'espèce *S. robustum*. C'est dans cette région que nous trouvons la plus grande diversité morphologique des clones *S. officinarum* (Daniels & Roach, 1987). Des analyses moléculaires avec des marqueurs RFLP ont montré que l'espèce *S. officinarum* était très proche de *S. robustum* (Lu *et al.*, 1994b). De plus, D'Hont *et al.*, (1998) ont montré que le nombre de chromosomes de base de *S. officinarum* est de  $x = 10$  comme *S. robustum* et que cette espèce est octoploïde comme un des cytotypes prédominants ( $2n = 80$ ) de *S. robustum*. Ces clones nobles auraient ensuite été diffusés vers la partie indomalaise (Inde et Chine) et vers les îles du Pacifique lors de migration humaine entre 1 500 et 1 000 ans avant notre ère. C'est à cette même période que les espèces *S. barberi* et *S. sinense* seraient apparues en Inde et en Chine respectivement. Les analyses moléculaires utilisant les marqueurs RFLP montrent que ces deux espèces sont issues d'une hybridation naturelle entre les espèces *S. spontaneum* et *S. officinarum* (Figure 1.3). Ces données ont été confirmées par des expériences de cytogénétique moléculaire (Genomic In Situ Hybridization - GISH) (D'Hont *et al.*, 2002) (Figure 1.4).

La canne à sucre a été propagée vers les pays occidentaux par les Perses vers l'an 500 de notre ère (Figure 1.2), puis par les Arabes qui ont développé la culture en Afrique et dans les îles de la méditerranée entre le VI<sup>ème</sup> et le VIII<sup>ème</sup> siècle. Au XV<sup>ème</sup> siècle, les Espagnols et les Portugais ont introduit la canne à sucre dans leurs possessions africaines. Christophe Colomb installe des plants de canne à sucre dès son second voyage dans les îles Caraïbes et vers 1505, ces îles exploitent le sucre de la canne pour la première fois. C'est le début d'une production prospère et lucrative. Au cours du XVI<sup>ème</sup> et début du XVII<sup>ème</sup> siècle, la culture de la canne



**Figure 1.3.** Scénario d'évolution et de domestication de la canne à sucre (Grivet *et al.*, 2006).

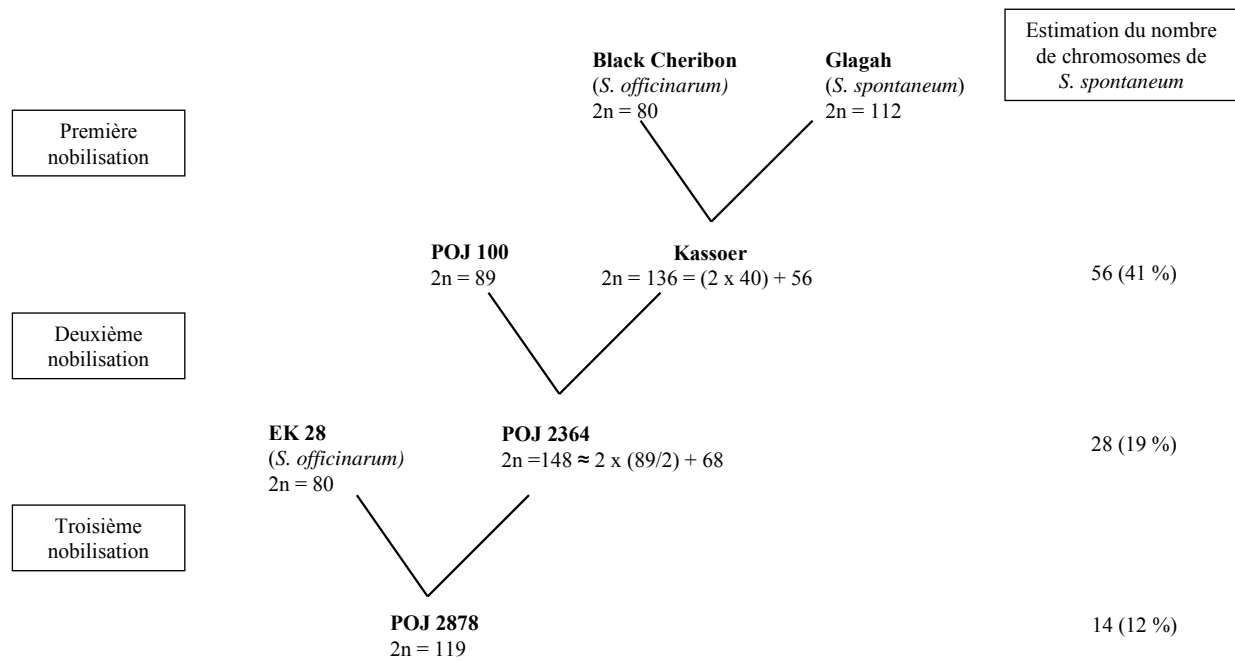


**Figure 1.4.** Hybridation génomique *in situ* (GISH) d'un clone de *S. barberi* ( $2n = 82$ ) en utilisant l'ADN génomique de *S. officinarum* (marqué en vert), l'ADN génomique de *S. spontaneum* (marqué en rouge) et l'ADNr 18S-26S (marqué en bleu). Les flèches blanches représentent les échanges de chromosomes interspécifiques (D'Hont *et al.*, 2002).

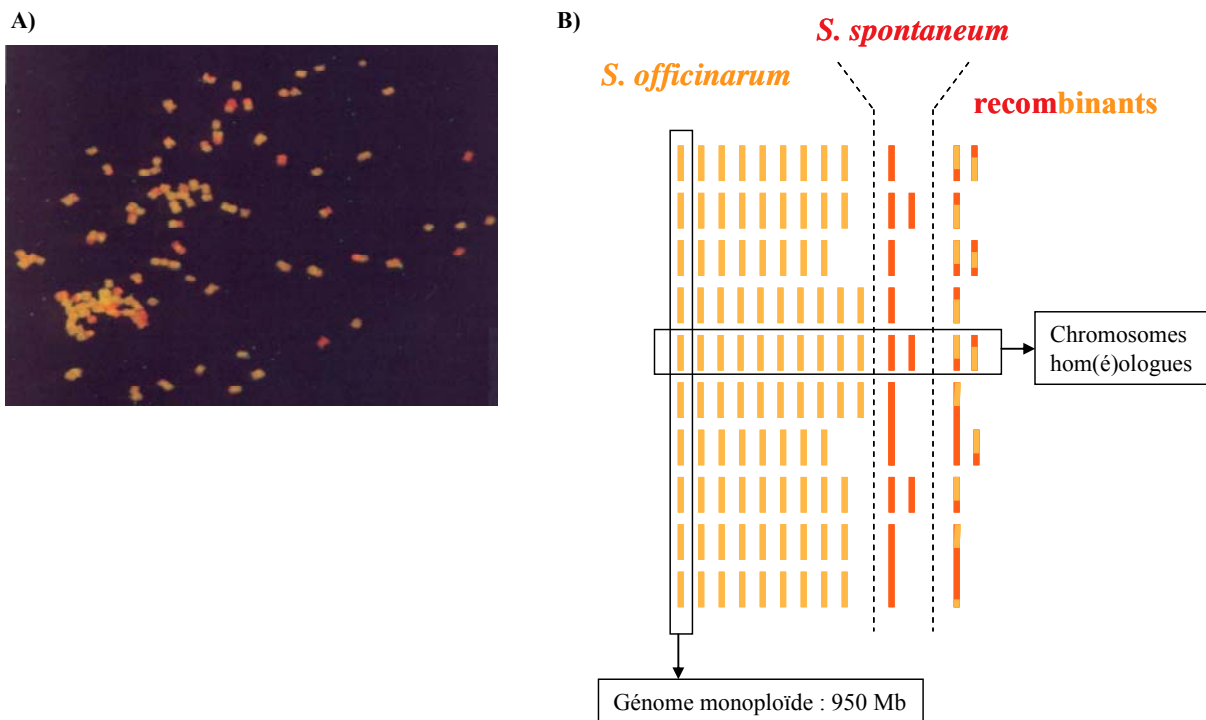
apparaît en Amérique du Sud (Mexique, Pérou et Brésil). A la moitié du XVIII<sup>ème</sup> siècle, la culture de la canne va s'intensifier avec son introduction aux Antilles (Guadeloupe et Martinique). L'industrie de la canne à sucre commence vers la fin du XVIII<sup>ème</sup> siècle avec l'utilisation de clones provenant de *S. officinarum* dits clones « nobles ». On pratiquait la substitution clonale qui consistait à remplacer des variétés nobles par des formes naturelles de la même espèce plus productives. Les clones les plus utilisés jusqu'en 1880 étaient des mâles stériles et on pensait que la canne à sucre se reproduisait seulement par multiplication végétative. Mais, vers le milieu du XIX<sup>e</sup> siècle, dans les îles de La Barbade, Stevenson (1965) a observé pour la première fois des plantules issues de graines.

Les premières variétés commerciales (POJ 100 et EK 28) issues de la multiplication sexuée de la canne à sucre ont été produites à partir de croisements de deux variétés nobles. Cependant, la variabilité génétique de *S. officinarum* ne permettait pas d'obtenir des variétés améliorées pour certains caractères agronomiques comme l'adaptation aux stress biotiques et abiotiques. Ainsi, d'autres espèces du genre *Saccharum* ont été introduites dans les programmes de sélection : *S. spontaneum* et *S. barberi*. Ces procédures d'introgression portent le nom de « nobilisation ». Elles consistent à croiser un clone noble (*S. officinarum*), riche en sucre, avec un clone d'une espèce parentée, puis de recroiser l'hybride avec un clone noble parental (backcross). Le clone noble n'est pas forcément le même à chaque génération de croisement, ceci afin de maintenir une vigueur élevée.

Les cultivars modernes sont tous issus de croisements entre ces premiers clones « nobilisés ». Selon Arcenaux (1965), la généalogie des cultivars modernes repose sur un nombre restreint de clones. Seuls 19 clones de *S. officinarum*, quelques clones de *S. spontaneum*, un clone de *S. barberi* et un clone de *S. robustum* auraient participé à l'élaboration des cultivars modernes. Ce rétrécissement potentiel de la variabilité en amont de la sélection des cultivars modernes a été considéré comme responsable du ralentissement du progrès génétique observé dans les années 1960. Mais une étude de diversité réalisée sur 109 variétés modernes et 53 clones de *S. officinarum* a montré que les cultivars modernes ont conservé une grande partie de la variabilité observée chez *S. officinarum* (Jannoo *et al.*, 1999).



**Figure 1.5. Transmission des chromosomes au cours de la nobilisation pour l'obtention du clone POJ 2878 (Bremer, 1961).**



**Figure 1.6. Structure du génome du cultivar R570 ( $2n = \text{ca } 115$ ).**

A) Hybridation génomique *in-situ* (GISH) de chromosomes du cultivar R570 en utilisant l'ADN génomique de *S. officinarum* (marqué en jaune) et l'ADN génomique de *S. spontaneum* (marqué en rouge) (D'Hont *et al.*, 1996).

B) Représentation schématique du génome de R570. Chaque trait vertical représente un chromosome. Les chromosomes issus du parent *S. officinarum* sont représentés en jaune et ceux du parent *S. spontaneum* en rouge. Les chromosomes présents sur une même ligne appartiennent au même groupe d'hom(é)ologie. Les chromosomes présents sur une même colonne appartiennent au même génome monoploïde.

### A.5. Structure des génomes des cultivars modernes

Les variétés cultivées actuellement sont dérivées de croisements entre *S. officinarum* et *S. spontaneum*. Ce sont des hybrides interspécifiques complexes et aneuploïdes avec un nombre de chromosomes variant de  $2n = 100$  à  $130$  (Sreenivasan *et al.*, 1987). Ce nombre élevé de chromosomes est la conséquence de la transmission chromosomique irrégulière lors des premières étapes de la nobilisation. Les clones *S. officinarum* utilisés comme parent femelle ont transmis la totalité de leurs chromosomes lors du premier croisement F1 et du premier rétrocroisement (Figure 1.5). Ce n'est qu'à partir du troisième rétrocroisement que *S. officinarum* transmettra  $n$  chromosomes (Bremer, 1961). Le mécanisme exact de la transmission  $2n + n$  n'a pas encore été parfaitement établi (Bhat & Gill, 1985). L'étape de la méiose est certainement perturbée car, malgré la formation de gamètes non réduits, des ségrégations de caractères maternels ont été observées parmi des hybrides F1 lors d'un croisement contrôlé de *S. officinarum* et *S. spontaneum* (Bremer, 1961). Deux hypothèses ont été avancées pour expliquer ce phénomène : (1) le doublement chromosomique (endoduplication) ou (2) la fusion de deux noyaux après la seconde division de la méiose. La nobilisation a donc provoqué une réduction rapide du nombre relatif de chromosomes de *S. spontaneum* à la méiose (Sreenivasan *et al.*, 1987) (Figure 1.5).

Des analyses d'hybridations *in-situ* (GISH) ont permis d'identifier l'origine spécifique des chromosomes de cultivars modernes de canne à sucre. D'Hont *et al.*, (1996) a montré que le génome d'un cultivar moderne réunionnais, R570, était constitué de 80 % de chromosomes issus de *S. officinarum*, de 10 % de chromosomes issus de *S. spontaneum* et de 10 % de chromosomes issus de recombinaisons entre des chromosomes de *S. officinarum* et *S. spontaneum* (Figure 1.6). Cuadrado *et al.*, (2004) a réalisé le même type d'analyses sur trois autres cultivars modernes. Leur génome est composé de 16 % de chromosomes issus de *S. spontaneum* et seulement de 5 % de chromosomes recombinants. Cette faible proportion du génome de *S. spontaneum* par rapport au génome de *S. officinarum* dans le génome des cultivars modernes est confirmée chez sept autres clones, avec seulement 15 à 27,5 % des chromosomes provenant de *S. spontaneum* et 8 à 13 % de recombinants (Piperidis *et al.*, 2010). Toutefois, des études utilisant des marqueurs moléculaires ont montré que le génome de *S. spontaneum* constituait la principale composante de la diversité des cultivars modernes (Lu *et al.*, 1994a; Jannoo *et al.*, 1999).

**Tableau 1.2. Cartes génétiques disponibles chez la canne à sucre.**

Clone ou cultivar	Nombre et types de marqueurs	Nombre de GC / GH*	Taille en cM	Référence
<i>S. spontaneum</i> (SES 208)	527 (RFLP, APPCR)	64 / 8	-	Al-Janabi <i>et al</i> , 1993 Silva <i>et al</i> , 1993 et 1995
<i>S. spontaneum</i> (IND81-146)	248 (RFLP)	69 / 10	2063	Ming <i>et al</i> , 1998
<i>S. spontaneum</i> (PIN84-1)	182 (RFLP)	69 / 7	1303	Ming <i>et al</i> , 1998
<i>S. officinarum</i> (LA Purple)	345 (RFLP, APPCR, AFLP)	74 / -	-	Guimaraes <i>et al</i> , 1999
<i>S. officinarum</i> (Green German)	270 (RFLP)	72 / 10	2034	Ming <i>et al</i> , 1998
<i>S. officinarum</i> Muntock Java	206 (RFLP)	72 / 9	1443	Ming <i>et al</i> , 1998
<i>S. robustum</i> Mol 5829	301 (RFLP, APPCR, AFLP)	65 / -	-	Guimaraes <i>et al</i> , 1999
Cultivar SP 701006	94 (RFLP)	25 / 8	-	D'Hont <i>et al</i> , 1994
Cultivar R570	408 (RFLP)	96 / 10	-	Grivet <i>et al</i> , 1996
	+322 (RFLP) +138 (AFLP) +138 (SSR)	162 / 8	-	données non publiées
Cultivar R570	1123 (AFLP, RLFP, SSR)	128 / 7	7675	Hoarau <i>et al</i> , 2001 Rossi <i>et al</i> , 2003 Raboin <i>et al</i> , 2006
Cultivar Q165	1074 (AFLP, SSR, RAF)	136 / 8	9058	Aitken <i>et al</i> , 2005
Cultivar MQ 76-53	536 (AFLP, RLFP, SSR)	105 / -	4300	Raboin <i>et al</i> , 2006

\* GC : Groupe de coségrégation; GH : Groupe d'homologie



Le génome du cultivar moderne R570 a été beaucoup étudié. Sa taille génétique est estimée à 17 000 cM pour une taille physique de 10 000 Mb (Hoarau *et al.*, 2001). La taille du génome monoploïde estimée est sensiblement égale à celle du sorgho et deux fois plus grandes que celle du riz : le génome monoploïde de *S. officinarum* est estimé à 960 Mb et celui de *S. spontaneum* à 750 Mb (D'Hont & Glaszmann, 2001). Chaque chromosome de base est potentiellement présent en dix ou douze exemplaires, avec des origines différentes. On parle alors de chromosomes homéologues.

#### A.6. Cartes génétiques de canne à sucre disponibles

La cartographie génétique est un outil performant pour l'analyse des génomes. Chez la canne à sucre, elle s'avère beaucoup plus difficile à mettre en œuvre que chez d'autres espèces du fait du degré élevé de ploïdie de son génome. Malgré la complexité de ce génome, plusieurs cartes génétiques d'espèces du genre *Saccharum* ont été développées (Tableau 1.2) :

- 3 cartes de *S. officinarum* (Ming *et al.*, 1998; Guimaraes *et al.*, 1999)
- 3 cartes de *S. spontaneum* (Aljanabi *et al.*, 1993; Silva *et al.*, 1993; Silva *et al.*, 1995; Ming *et al.*, 1998)
- 1 carte de *S. robustum* (Guimaraes *et al.*, 1999)

Des travaux de cartographie génétique ont aussi été menés sur les cultivars modernes (D'Hont *et al.*, 1994; Grivet *et al.*, 1996; Hoarau *et al.*, 2001; Rossi *et al.*, 2003; Aitken *et al.*, 2005; Raboin *et al.*, 2006) (Tableau 1.2) mais aucune carte génétique saturée du génome de la canne à sucre n'est disponible. Les deux cartes les plus complètes concernent deux cultivars modernes : Q 165, qui compte 136 groupes de coségrégation rassemblés en 8 groupes d'homologie (Aitken *et al.*, 2005) et R570, qui compte 128 groupes de coségrégation rassemblés en 7 groupes d'homologie (Hoarau *et al.*, 2001; Rossi *et al.*, 2003; Raboin *et al.*, 2006). La longueur cumulée des cartes est de 9 058 cM et 7 675 cM respectivement. La plupart des cartes génétiques de canne à sucre ont été réalisées avec des marqueurs RFLP spécifiques des locus qu'ils révèlent. Cette caractéristique des marqueurs RFLP permet de réunir les groupes de coségrégation (GC) en groupes de liaison sur la base de sondes en commun. Chaque GC correspond à un chromosome ou à un segment de chromosome lorsque

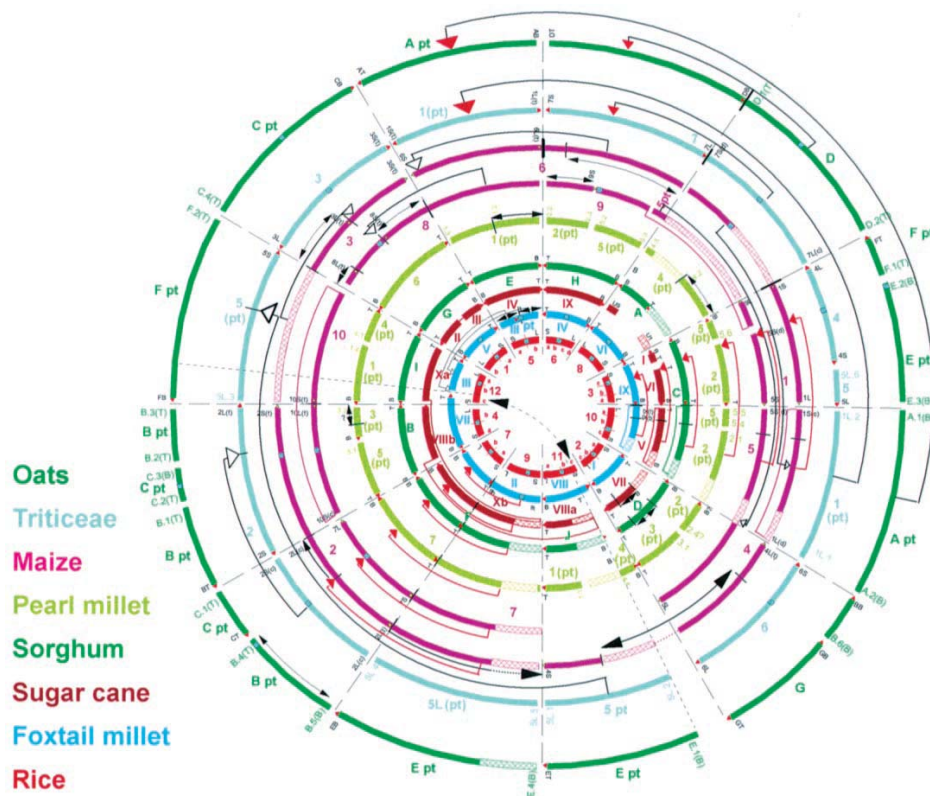


Figure 1.7. Synténie entre les principaux génomes de céréales (Gale & Devos, 1998).

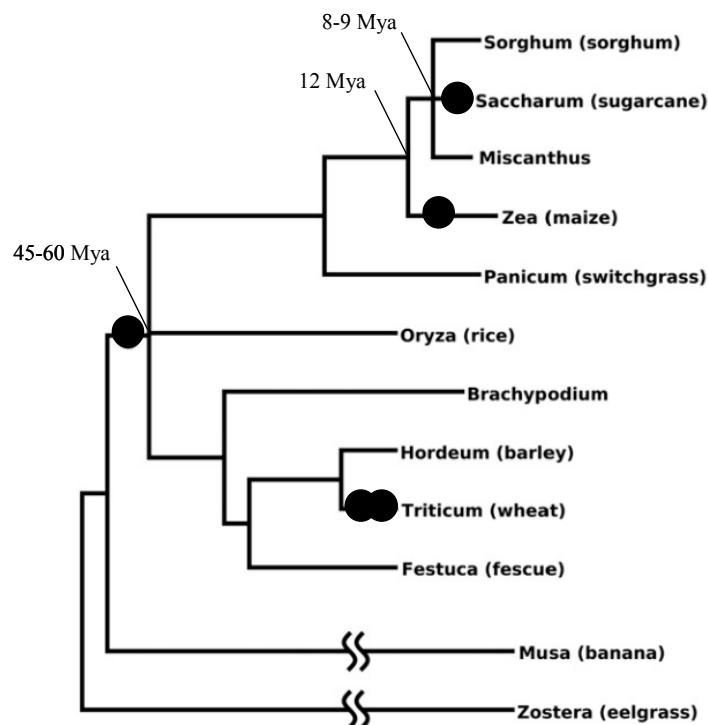


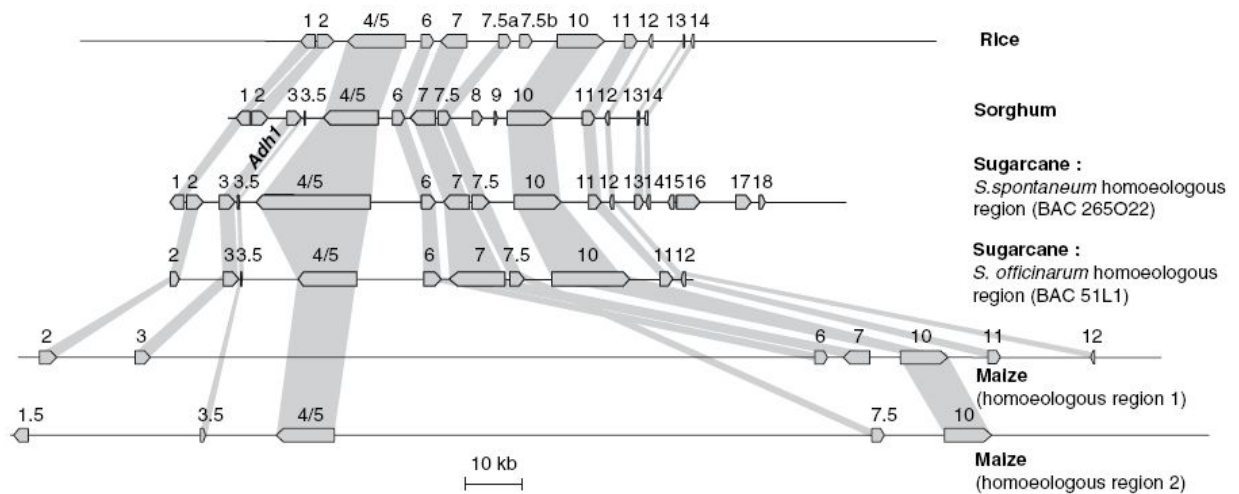
Figure 1.8. Relations phylogénétiques entre les principales espèces de graminées (d'après Paterson *et al*, 2009). L'arbre phylogénétique est enraciné avec deux espèces "outgroup" (*Musa* et *Zostera*). Les événements de duplication de génome entier (polyploïdisation) sont indiqués par des ronds noirs. Les temps de divergence sont estimés en millions d'années (Mya).

la carte n'est pas saturée, et les groupes de liaison correspondent aux groupes d'homologie. Les marqueurs de type SSR (Single Sequence Repeats) ont aussi été utilisés chez la canne à sucre (Rafalski & Tingey, 1993). Ce type de marqueurs codominants et simples d'utilisation (amplification PCR, Polymerase Chain Reaction) permettent, comme les RFLP, de réaliser des ponts entre groupes de liaison, voire entre différentes cartes. Les marqueurs AFLP (Amplified Fragment Length Polymorphism) sont aussi très utilisés en complément des marqueurs RFLP et SSR. Ce sont des marqueurs qui permettent de générer un nombre important de marqueurs par couples d'amorces.

#### **A.7. Relations synténiques entre génomes de Poacées**

De nombreux travaux montrent un degré élevé de conservation des génomes entre les différentes espèces de graminées (Figure 1.7) (Feeling, 2001; Bolot *et al.*, 2009). Cette forte conservation de l'organisation des chromosomes permet de transférer, par comparaison des cartes génétiques, les données acquises sur une espèce particulière aux autres espèces de graminées et ainsi d'identifier des régions/gènes d'intérêt (Gale & Devos, 1998). Les relations synténiques entre la canne à sucre, le sorgho, le maïs et le riz ont été étudiées. Une première analyse a été réalisée en cartographiant des sondes de maïs sur le génome de la canne à sucre (D'Hont *et al.*, 1994). Une relative conservation de la synténie entre ces deux espèces a été mise en évidence, relation seulement perturbée par l'origine tétraploïde du maïs (Grivet *et al.*, 1994). Le développement de cartes génétiques chez le sorgho a permis de comparer l'organisation du génome de cette espèce avec celui de la canne à sucre. Les comparaisons de génomes ou de régions de génome ont révélé un degré élevé de conservation de la synténie entre ces deux espèces (Dufour *et al.*, 1997; Glaszmann *et al.*, 1997; Guimaraes *et al.*, 1997). Les relations synténiques avec le génome du riz ont été étudiées dans un premier temps par l'intermédiaire de cartes génétiques de maïs et de sorgho puis par comparaison directe entre la canne à sucre et le sorgho (Glaszmann *et al.*, 1997; Asnaghi *et al.*, 2000). Une bonne synténie entre les génomes de la canne à sucre et du riz a été identifiée, malgré une distance phylogénétique plus importante qu'entre la canne à sucre et le sorgho ou le maïs (Figure 1.8).

Avec l'arrivée des nouvelles technologies de séquençage, les études de synténie se sont étendues à l'analyse de la microcolinéarité avec les Graminées modèles, afin de comparer les dynamiques d'évolution de ces espèces. Les génomes du sorgho, de



**Figure 1.9. Conservation des gènes entre des régions orthologues de la canne à sucre, du sorgho, du riz et deux régions hom(é)ologues du maïs (d'après Jannoo *et al*, 2007).**  
 Les gènes sont représentés par des flèches pleines grises. Les aires grisées connectées représentent la conservation entre les gènes orthologues.

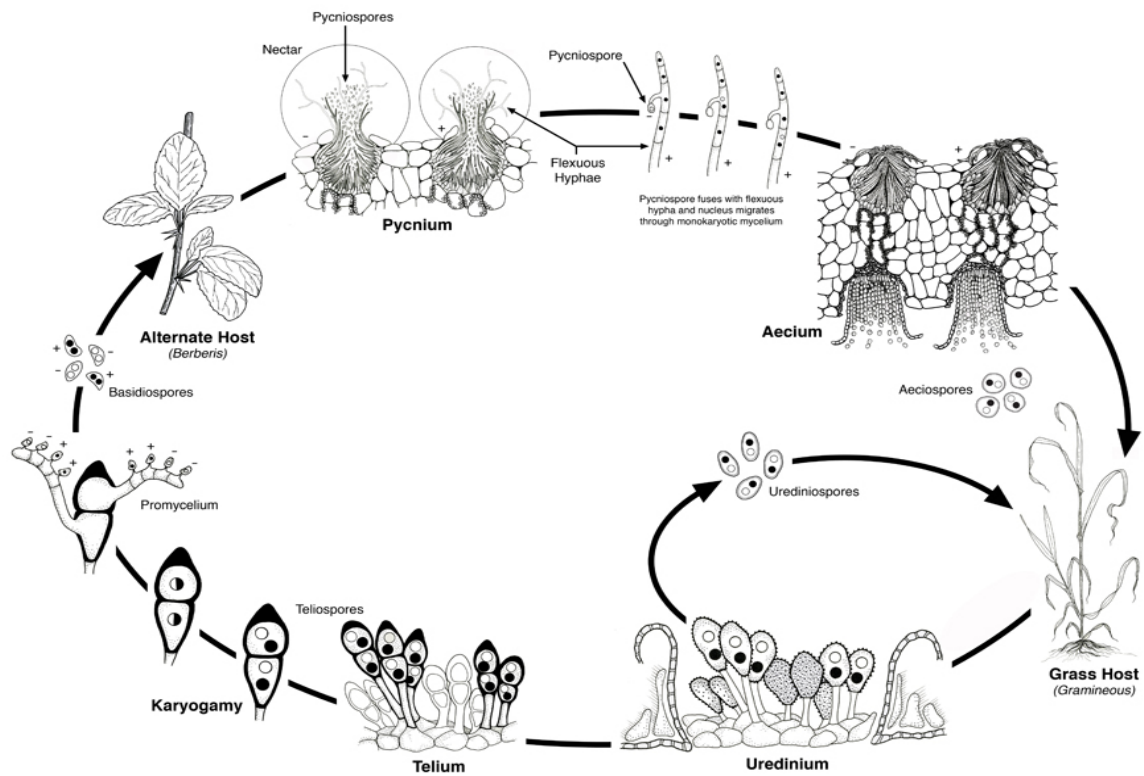
*Brachypodium distachyon*, du maïs et de deux variétés de riz (Nipponbare et 93-11) ont été complètement séquencés (Goff *et al.*, 2002; Yu *et al.*, 2002; Paterson *et al.*, 2009; Schnable *et al.*, 2009; Vogel *et al.*, 2010). Toutes ces données disponibles permettent d'augmenter la profondeur des analyses de génomique comparative et de faciliter les démarches de cartographie comparée et de recherche de gènes/régions d'intérêt.

Ce type d'étude de micro-colinéarité au sein des Poacées a par exemple été réalisé au niveau d'une région comprenant un gène codant pour une alcool deshydrogénase Adh1. Une première comparaison avait été faite entre les génomes du riz, du sorgho et du maïs ; une très bonne colinéarité de la région Adh1 avait été observée (Ilic *et al.*, 2003). Pour la canne à sucre, deux séquences homéologues (de 97 kb et 126 kb), l'une dérivée de *S. officinarum* et l'autre de *S. spontaneum*, co-existant chez le cultivar moderne R570 et correspondant à la région Adh1, ont été séquencées, annotées finement et analysées. Les deux haplotypes montrent une parfaite colinéarité au niveau des gènes et une très forte conservation de structure, aussi bien entre eux qu'avec les séquences orthologues du riz et du sorgho (Figure 1.9). Une telle conservation de séquence entre canne/sorgho et canne/riz, qui ont pourtant divergé respectivement il y a 7 à 8 millions d'années et 50 millions d'années, peut être reliée à la stabilité génomique de cette région (Jannoo *et al.*, 2007).

## **B. Les champignons de la rouille**

### **B.1. Caractéristiques générales et taxonomie**

Environ 7 000 espèces de champignons, réparties en 100 genres, sont regroupées sous le nom de rouille en référence à l'aspect et à la couleur des symptômes qu'elles provoquent sur les plantes (Cummins & Hiratsuka, 2003). Les rouilles sont des champignons pathogènes d'un très grand nombre d'espèces végétales (dicotylédones ou monocotylédones). Ils appartiennent à la classe des Basidiomycètes et à l'ordre des Urédinales. Les champignons de ce groupe se développent exclusivement sur les tissus vivants des plantes et leur cycle de vie est complexe avec la production de cinq types de spores différentes (Swann *et al.*, 2001) : basidiospores, pycniospores, aecidiospores, urédospores et téléutospores. Le cycle de développement complet des Urédinales peut s'effectuer soit sur des plantes de la même



**Figure 1.10.** Cycle de vie d'une rouille de blé : *Puccinia graminis* (Leonard & Szabo, 2005)

espèce (parasites autoïques) soit sur des plantes de différentes espèces (parasites hétéroïques) (Figure 1.10). Une autre caractéristique importante des champignons de la rouille est leur spécificité d'hôte. Chaque groupe de champignon n'est capable d'infecter qu'une certaine gamme de taxons végétaux (Cummins & Hiratsuka, 2003).

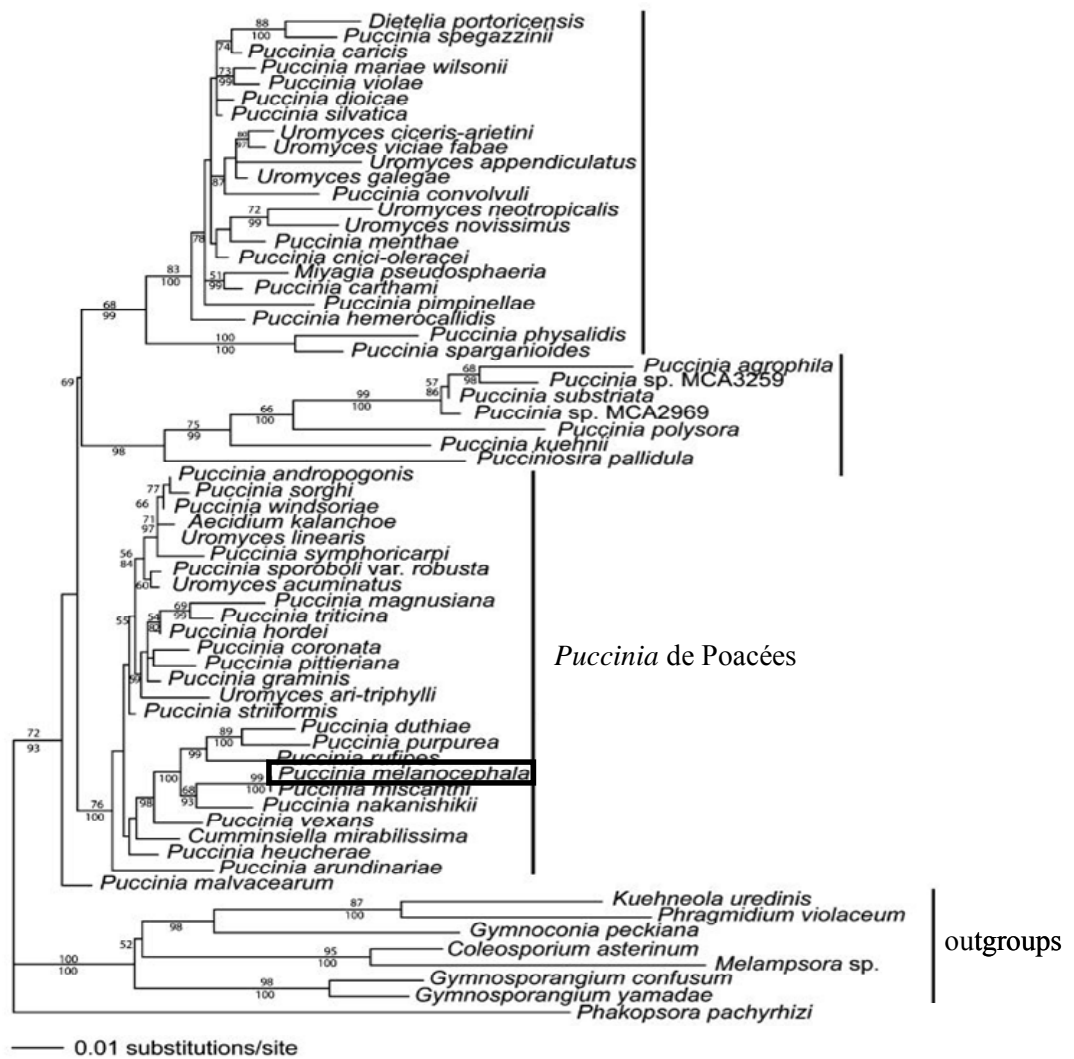
Le nombre de genres et d'espèces étant très important, la classification des champignons de la rouille a longtemps évolué. Traditionnellement, elle était basée sur la morphologie des téliospores. Ainsi, deux familles ont été acceptées par Dietel (1928) : la famille des Pucciniaceae (téliospores dit pétiolées) et la famille des Melampsoraceae (téliospores dit sessiles). Néanmoins, Cummins et Hiratsuka (1983) ont montré l'importance d'autres caractères que la morphologie des téliospores : la disposition des spores des urédospores, la morphologie des urédies ou la morphologie des spermogonies. Les auteurs ont donc proposé une nouvelle classification regroupant 12 familles. Cette classification est toujours d'actualité et a été appuyée par des études moléculaires de phylogénie (Maier *et al.*, 2003) (Figure 1.11).

Dans la famille des Pucciniaceae, le genre *Puccinia* rassemble entre 3 000 et 4 000 espèces parmi lesquelles certaines sont responsables de redoutables maladies chez les Poacées comme la rouille noire des céréales (*P. graminis*), la rouille jaune du blé (*P. striiformis*), la rouille commune du maïs (*P. sorghi*) ou la rouille brune de la canne à sucre (*P. melanocephala*). Des études de phylogénie montrent que les champignons responsables des rouilles chez les Poacées forment un clade bien distinct dans le genre *Puccinia* (Van der Merwe *et al.*, 2006; Dixon *et al.*, 2010)

## **B.2. La rouille brune chez la canne à sucre : *Puccinia melanocephala***

### **B.2.1 Historique et distribution**

Il existe deux types de rouilles chez la canne à sucre : la rouille brune et la rouille orange dont les agents pathogènes sont *P. melanocephala* et *P. kuehnii* respectivement. Bien que les spores de ces deux pathogènes aient une morphologie qui permette de les différencier, ils ont souvent été confondus car les symptômes qu'ils provoquent sur les feuilles sont très semblables (Virtudazo *et al.*, 2001). La répartition géographique de *P. melanocephala* s'étend aujourd'hui dans pratiquement tous les pays producteurs de canne à sucre. *P. kuehnii* quant à



**Figure 1.11. Arbre phylogénétique d'un échantillon de champignons appartenant au genre *Puccinia*.**

L'encadré noir correspond au champignon responsable de la rouille brune chez la canne à sucre : *P. melanocephala* (d'après Dixon *et al*, 2010).



lui, a été décrit en Asie et en Australie. Des rapports récents ont confirmé la présence du champignon en Amérique du Nord et en Amérique du Sud (Comstock *et al.*, 2008; Flores *et al.*, 2009).

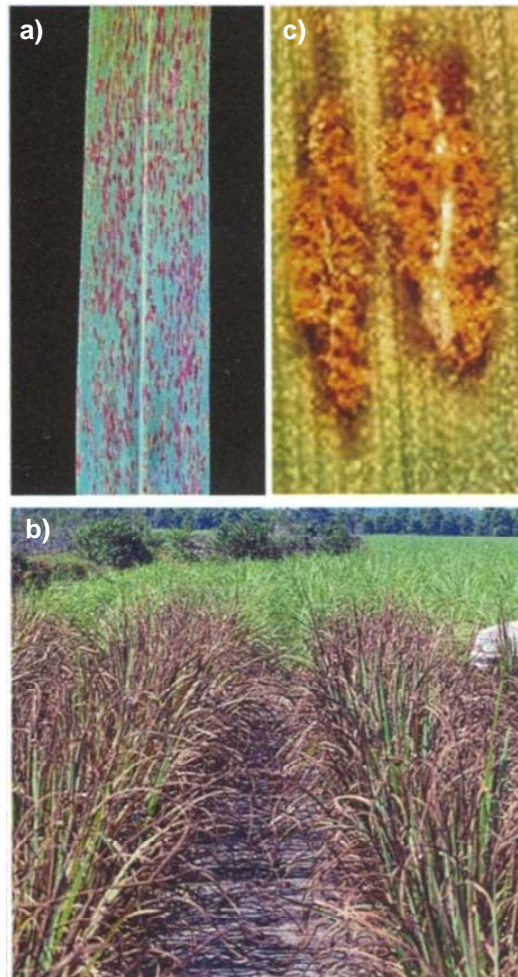
La rouille brune causée par le champignon *P. melanocephala* a été signalée pour la première fois en Inde dans les années 1950 quand une épidémie s'est développée sur un clone de canne à sucre cultivé depuis quelques années (Ryan & Egan, 1989). La maladie ne gagne une importance mondiale qu'en 1978 quand elle est signalée pour la première fois en Australie et dans la région des Caraïbes puis l'Amérique du Nord et le Mexique (Comstock & Raid, 1994). La maladie s'est propagée très vite et a causé de nombreuses pertes économiques (Comstock *et al.*, 1992b). Purdy *et al.*, (1983) ont enregistré au Mexique des pertes de 1,5 à 21,7 tonnes par hectare entre 1978 et 1979 pour le cultivar B4362, qui occupait 28 % des sols pour la production de canne à sucre. L'industrie sucrière cubaine a estimé une perte de un million de tonnes de sucre en 1979.

### **B.2.2 Races de *P. melanocephala***

Lors de l'épidémie de rouille brune en Inde en 1950, il a été noté que certains cultivars avaient des niveaux de résistance différents suivant la région où ils étaient cultivés (Srinivasan & Muthaiyan, 1965). Dean et Purdy (1984) décrivent trois cultivars dont la réaction vis-à-vis de la rouille est subitement passée de résistante à sensible dans certaines parcelles. Après avoir écarté les possibilités de variation des conditions environnementales entre les parcelles, les auteurs ont interprété ce phénomène comme une variation du niveau de résistance des cultivars et donc l'existence de différentes races de *P. melanocephala* avec des pouvoirs pathogènes plus ou moins importants.

Plusieurs études suggèrent la présence de différentes races en comparant les niveaux de résistance d'un cultivar résistant. Afin de se soustraire des facteurs environnementaux, les études citées ci-dessus ont été réalisées en conditions contrôlées :

- Six isolats collectés en Inde ont provoqué des réactions différentes sur 18 cultivars (Srinivasan & Muthaiyan, 1965).
- Deux isolats collectés à Hawaï ont provoqué une réaction différente sur le cultivar H65-7052 (Comstock, 1985).



**Figure 1.12. Exemples de symptômes causés par l'agent de la rouille brune *P. melanocephala* chez la canne à sucre.**

a) symptômes sur feuille de canne à sucre et b) en champ. c) pustule de *P. melanocephala* sur feuille de canne à sucre.

- Cinq isolats collectés en Floride ont donné une réaction différente sur six cultivars (Comstock & Raid, 1994). Shine *et al*, (2005) ont réalisé le même type d'étude avec ces cinq isolats. Les auteurs suggèrent qu'il existe au moins trois races différentes en Floride.

Par contre, Taylor (1992), a utilisé une collection d'urédospores de *P. melanocephala* récoltées entre 1978 et 1986 (15 échantillons) en Australie et les ont testés sur 12 cultivars résistants. Aucune réaction différentielle n'a été observée sur les 12 cultivars testés. Les auteurs ont déduit qu'il existait qu'une seule race physiologique de *P. melanocephala* en Australie.

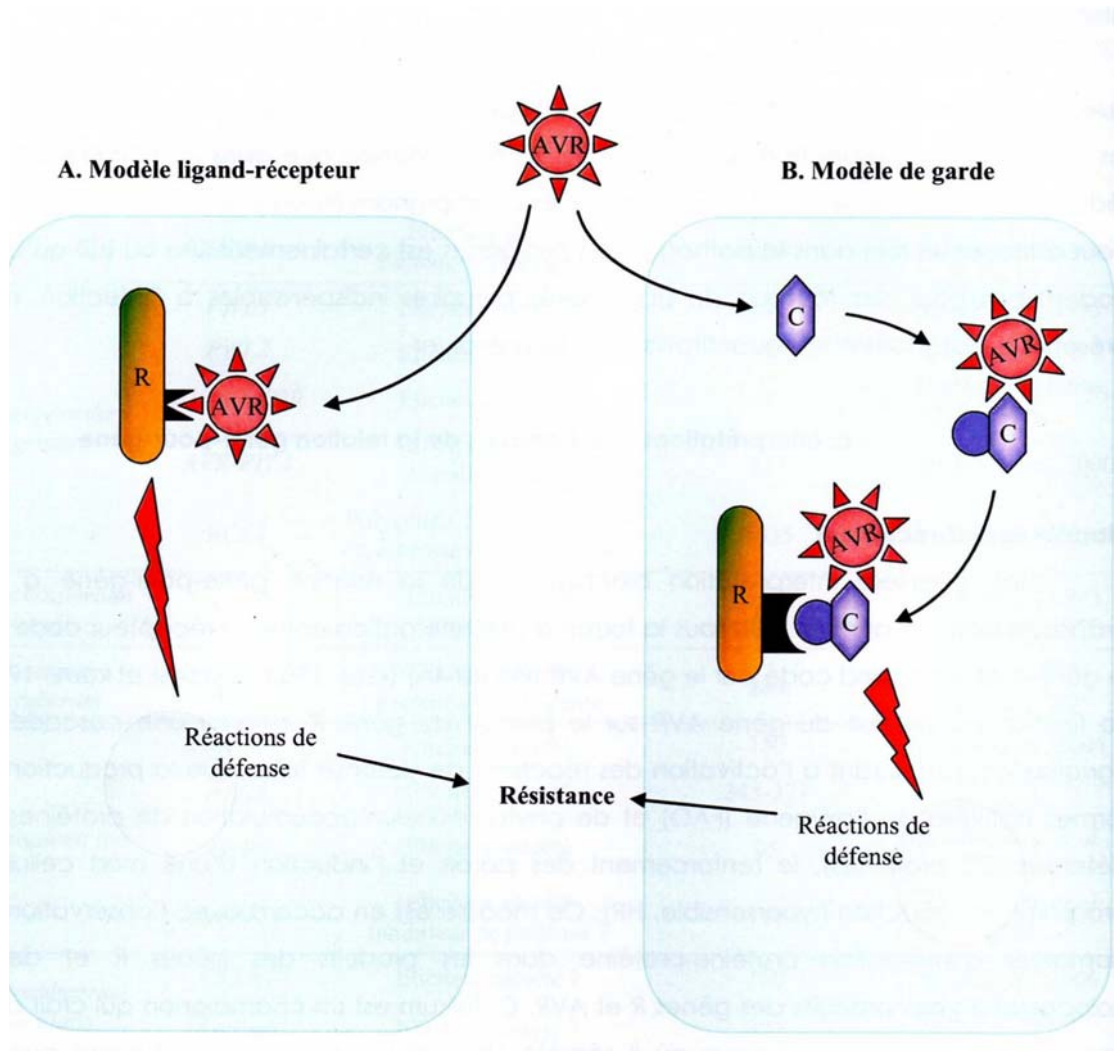
### **B.2.3 Symptômes et transmission de la rouille brune**

Le développement de la maladie commence par l'apparition de tâches jaunâtres visibles sur les deux faces de la feuille. Ces tâches s'allongent, deviennent marron à marron-rouge, et un halo chlorotique se forme autour de la lésion (Figure 1.12). Des pustules contenant les urédospores (de 2 à 20 mm de long pour 1 à 3 mm de large) se développent sur la face inférieure de la feuille. Sur les feuilles les plus sévèrement touchées, les pustules coalescent et forment de grandes surfaces nécrotiques. A maturité, les pustules produisent de nombreuses urédospores de couleur marron-orangé. Ces urédospores se déposent sur d'autres feuilles et germent pour coloniser les tissus de la feuille. Douze à 14 jours plus tard, d'autres urédospores seront formées et iront coloniser une autre feuille.

Les urédospores sont essentiellement propagées par le vent et, dans une moindre mesure, par l'eau et l'homme, entre parcelles de canne à sucre par exemple. Le mode de transmission de cette maladie expliquerait son aptitude à se propager rapidement.

Le spectre d'infection de *P. melanocephala* est principalement constitué des différentes espèces du genre *Saccharum* spp. mais certaines lésions ont déjà été observées chez *Erianthus nolvus* et *Narenga porphyrocoma* (Ryan & Egan, 1989).

Les symptômes de la maladie sont évalués visuellement grâce à une échelle de notation graduelle inspirée de celle proposée par Tai *et al*, (1981), allant de 1 à 9. Les individus dont la note est inférieure à 2 sont résistants (absence de pustules). Les individus notés de 2 à 9 sont sensibles à des degrés différents. Un individu noté 2 présente quelques pustules tandis qu'un individu noté 9 présente un très grand nombre de pustules.



**Figure 1.13. Modèles biochimiques de la relation gène-pour-gène (Dangl & Jones, 2001).**

A. Modèle ligand-récepteur. La protéine de résistance R se fixe sur la protéine d'avirulence AVR. L'interaction entre ces deux protéines déclenche la cascade de signalisation aboutissant à l'activation des gènes de défenses.

B. Modèle de garde. La protéine AVR se fixe sur une protéine cible (C), interaction qui induit un changement de conformation. Ce changement de conformation est détecté par la protéine R et conduit à l'activation des gènes de défenses.

Il existe deux moyens pour lutter contre la rouille brune de la canne à sucre :

- l'utilisation de fongicides. Cette méthode est très efficace mais reste très onéreuse sur les grandes superficies de culture de la canne à sucre et est donc très peu utilisée.
- l'utilisation de cultivars résistants est le moyen le plus utilisé. Le développement rapide de cultivars résistants qui a permis d'enrayer la première épidémie de rouille est essentiellement dû à la forte héritabilité du caractère rouille (Taylor *et al.*, 1982; Comstock *et al.*, 1992a; Daugrois *et al.*, 1996).

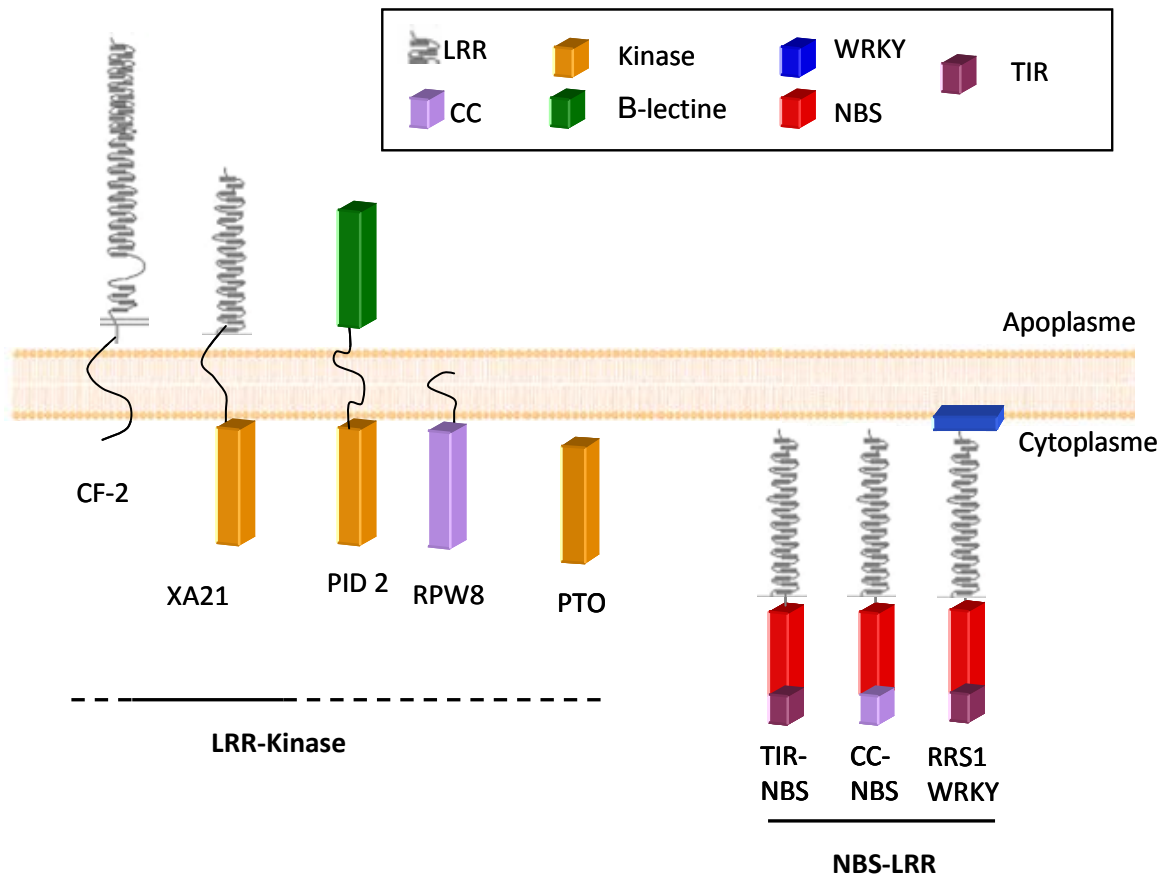
## **C. Les interactions plantes / microorganismes phytopathogènes**

### **C.1 Mécanismes de résistance**

Flor (1971) a décrit l'existence de résistances spécifiques monogéniques dans l'interaction lin / *Melampsora lini* et a énoncé la théorie dite gène-pour-gène, qui permet d'expliquer la grande majorité des résistances monogéniques dominantes. Dans ce type d'interaction, les mécanismes de résistance sont induits si et seulement si la protéine codée chez la plante par un gène de résistance (gène *R*) reconnaît la protéine codée chez l'agent pathogène par un gène dominant dit « d'avirulence » (gène *Avr*). On parle alors d'interaction incompatible, la plante est résistance et le pathogène avirulent. Lors d'une réaction incompatible, la réaction de défense de la plante implique une reconnaissance précoce des protéines d'avirulence de l'agent pathogène par les produits du gène de résistance. Ensuite, une cascade d'évènements va entraîner la transduction du signal. Ce signal entraînera la mise en place des mécanismes de défense, dont la réaction hypersensible qui entraîne la mort cellulaire programmée des cellules infectées ainsi que la mort de l'agent pathogène.

D'un point de vue biochimique, il existe deux modes de reconnaissance de la protéine d'avirulence par la protéine de résistance (Figure 1.13) :

- le modèle ligand-récepteur (Staskawicz *et al.*, 1995) : le récepteur est codé par le gène *R* et le ligand est codé par le gène *Avr*. Dans ce modèle, la protéine d'avirulence est



**Figure 1.14. Les grandes classes de protéines codées par les gènes de résistance *R* chez les plantes (Hammond-Kosack & Parker, 2003).**

reconnue directement par la protéine de résistance. L'interaction directe entre les deux protéines déclenche une cascade de signalisation conduisant à la réaction de défense.

- le modèle de garde (Dangl & Jones, 2001): les produits des gènes de résistances ne reconnaissent pas directement les produits des gènes d'avirulence. Les protéines *R* reconnaissent un complexe protéine de garde / protéine d'avirulence. Dans le modèle le plus simple, la fixation de la protéine d'avirulence sur la protéine de garde induit un changement de conformation du complexe qui déclenche la cascade de signalisation. Par exemple, l'étude de l'interaction *Arabidopsis thaliana* / *Pseudomonas syringae* a révélé que la protéine de résistance RPM1 n'interagit pas directement avec la protéine d'avirulence AVRPM1 mais aurait besoin de la protéine de garde RIN4 (Mackey *et al.*, 2002)

## C.2 Les gènes *R* de résistance

Plus d'une cinquantaine de gènes *R* conférant une résistance à des virus, bactéries, champignons et nématodes, a été caractérisée au niveau moléculaire chez différentes espèces végétales (van Ooijen *et al.*, 2007). La plupart des protéines codées par les gènes *R* possèdent deux domaines caractéristiques : un domaine pour la reconnaissance de l'agent pathogène et un domaine pour la transduction du signal. En fonction des motifs protéiques très conservés dans ces deux types de domaines, les gènes de résistance ont été classés en deux groupes majeurs (Figure 1.14) :

### - Les NBS-LRR

Cette classe de gènes de résistance est la plus répandue dans le monde végétal. Les protéines codées par ces gènes sont composées : (i) d'un domaine LRR (Leucine Rich Repeat), côté C-terminal, qui est un domaine d'interaction protéine-protéine qui joue un rôle dans la reconnaissance de l'agent pathogène (Dodds *et al.*, 2006) (ii) d'un domaine NBS (Nucleotide Binding Site), côté N-terminal, impliqué dans la transduction du signal (Tameling *et al.*, 2006). Ces protéines sont généralement intracellulaires. D'autres motifs protéiques, ayant également un rôle dans la signalisation, peuvent venir s'ajouter au domaine NBS-LRR : soit des





domaines de type CC (Coiled Coiled) ou LZ (Leucine Zipper) sur la partie C-terminale, soit un domaine TIR (Toll and Interleukin-1 Receptors). Les gènes *Rpp1*, *Rpp25* et *Rpp5* dans l'interaction *Arabidopsis thaliana* / *Peronospora parasitica* en sont des exemples.

Aucun gène codant pour une protéine avec un domaine TIR n'a été identifié chez les monocotylédones. Nous pouvons trouver un facteur de transcription de type WRKY en N-terminale. C'est le cas du gène *Rrs1r* dans l'interaction *Arabidopsis thaliana* / *Ralstonia solanacearum* (Deslandes *et al.*, 2002).

#### - Autres types de gènes de résistance

Comme les NBS-LRR, les LRR-kinases possèdent deux domaines : un domaine pour la reconnaissance (LRR) et un autre pour la signalisation (domaine kinase). Parmi les gènes de résistance clonés, ce groupe est le plus diversifié au niveau de la structure des protéines codées. Une majorité de ces protéines présente un domaine transmembranaire (TM) qui sert d'ancrage dans la membrane. Les LRR-TM-Sérine/Thréonine kinases sont codées par plusieurs gènes bien caractérisés comme *Xa21* dans l'interaction riz / *Xanthomonas oryzae* (Song *et al.*, 1995). Les Beta-lectine-TM-Sérine/Thréonine Kinases (le domaine LRR est remplacé par une  $\beta$ -lectine) sont codées par *Pid2* dans l'interaction Riz/*Magnaporthe oryzae* (Chen *et al.*, 2006).

Les sous-classes de protéines codées par des gènes *R* présentent une absence d'un des deux domaines. Les LRR-TM n'ont été identifiés que dans l'interaction tomate / *Cladosporium fulvum* (*Cf4*, *Cf5*, *Cf9*) (Joosten & de Wit, 1999). Cette sous-classe de gènes ne comporte pas de kinase. Les Sérine /Thréonine kinases (S/T kinase) ne comportent que des domaines kinases intracellulaires. Le gène *Pto* dans l'interaction Tomate / *Pseudomonas syringae* (un domaine kinase) et le gène *Rpg1* dans l'interaction orge / *Puccinia graminis* (deux domaines kinases en tandem) codent pour des protéines appartenant à cette famille (Martin *et al.*, 1993; Brueggeman *et al.*, 2002). Dans ces deux classes de gène de résistance, il est possible que d'autres gènes remplissent la fonction du domaine manquant.

### C.3 Gènes de résistance à la rouille chez les Poacées



La rouille peut engendrer de très grandes pertes économiques chez les espèces végétales de la famille des Poacées, en causant trois types de maladie :

- rouille jaune (« stripe rust »)
- rouille brune (« leaf rust »)
- rouille noire (« stem rust »).

De nombreuses études ont permis d'identifier au niveau génétique les gènes de résistance à la rouille chez les Poacées mais très peu de gènes ont été caractérisés au niveau moléculaire. Le blé reste l'espèce la plus étudiée pour cette interaction. Pour chaque type de maladie, de nombreux gènes de résistance aux différentes rouilles ont été identifiés chez cette espèce :

- les gènes de résistance *Yr* dans l'interaction *Triticum aestivum* / *P. striiformis* (rouille jaune). Actuellement, 70 gènes de résistance à ce pathogène ont été identifiés suivant le génotype de blé utilisé et la race du champignon (Chen, 2005). Seul le gène *Yr36* a été cloné (Fu *et al.*, 2009). Il s'agit d'une S/T kinase avec un domaine START qui permet de réaliser des échanges de lipide ou de stérol (Alpy *et al.*, 2009). L'action de cette kinase-START est dépendante de la température : à une température entre 25-35°C, le gène confère la résistance à la plante mais à une température inférieure à 15°C, la plante est sensible (Fu *et al.*, 2009).

- les gènes de résistance *Lr* dans l'interaction *Triticum aestivum* / *P. triticina* (rouille brune). Plus de 50 gènes ont été identifiés dans le génome du blé et ont été utilisés dans les programmes de sélection (Feuillet *et al.*, 2003). Seulement quatre gènes ont pu être caractérisés au niveau moléculaire. Trois d'entre eux codent pour des protéines de type CC-NBS LRR : *Lr1*, *Lr21* et *Lr10* (Feuillet *et al.*, 2003; Huang *et al.*, 2003; Qiu *et al.*, 2007). Un autre gène cloné récemment confère une résistance partielle depuis plus de 50 ans à trois maladies (rouille brune, rouille jaune et oïdium). Ce gène se nomme *Lr34* et il appartient à la famille des transporteurs ABC qui n'appartient pas à une famille classique de gènes de résistance (Krattinger *et al.*, 2009).

- les gènes de résistance *Sr* dans l'interaction *Triticum aestivum* / *P. graminis* (rouille noire). Quarante-six gènes de résistance ont été identifiés mais aucun n'a été caractérisé au niveau moléculaire (Ayliffe *et al.*, 2008).



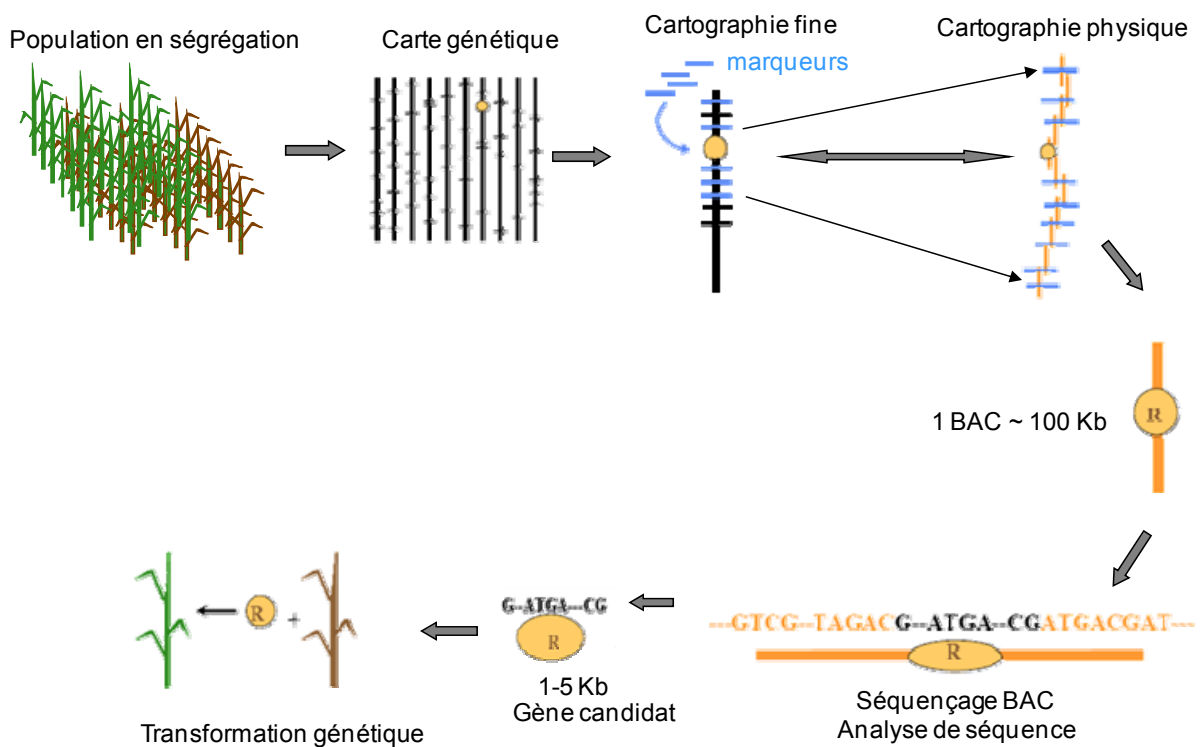
Selon Chelkowski *et al.*, (2003), 27 gènes ont été identifiés dans l'interaction orge / rouille : 6 gènes de résistance *Rpg* pour la rouille noire (*P. graminis*), 4 gènes de résistance *Rps* pour la rouille jaune (*P. striiformis*) et 17 gènes *Rph* pour la rouille brune (*P. hordei*). Seulement deux gènes de résistance impliqués dans l'interaction orge / *P. graminis* ont été clonés. *Rpg1* a été le premier gène cloné par Brueggeman *et al.*, (2002). Le produit du gène est constitué de deux S/T kinases en tandem. Récemment, le gène *Rpg5* a lui aussi été caractérisé au niveau moléculaire (Brueggeman *et al.*, 2008). La protéine de résistance comprend un domaine NBS, un domaine LRR et un domaine S/T kinase avec la présence de deux domaines transmembranaires, une structure assez atypique pour un gène de résistance.

Un seul autre gène de résistance a été cloné dans une interaction entre une Poacée et la rouille. Il s'agit du gène *Rp1D* du maïs pour la résistance à *P. sorghi* (Collins *et al.*, 1999). La structure de la protéine Rp1D correspond à une NBS-LRR. D'autres gènes ont été identifiés dans l'interaction maïs / *P. polysora* comme *Rpp9* (Holland *et al.*, 1998) et *RppQ* (Zhou *et al.*, 2007). Les gènes de résistance *Pg* interviennent dans l'interaction avoine / *P. graminis* (Harder, 1999). L'expression du gène *Pg-a*, comme celle du gène *Yr36*, est température-dépendante (Adhikari *et al.*, 2000). Mais aucun de ces gènes de résistance n'a été cloné.

La majorité de ces gènes a été caractérisée par clonage positionnel. Chez la canne à sucre, deux gènes ont été identifiés comme gènes de résistance à la rouille brune (*P. melanocephala*) : *Bru1* chez le cultivar réunionnais R570 et *Bru2* chez le clone MQ 76-53 (Daugrois *et al.*, 1996; Raboin *et al.*, 2006). *Bru1* confère une résistance durable depuis plus de 30 ans dans l'ensemble des zones où R570 est cultivé. Seules des données de cartographie sont disponibles pour le gène *Bru2* (Raboin *et al.*, 2006).

### **D. Clonage positionnel chez les plantes à grand génome**

#### **D.1. Contraintes du clonage positionnel chez les plantes à grand génome**



**Figure 1.15. Schématisation des étapes du clonage positionnel d'un gène d'intérêt.**

Pour cloner un gène dont on ne connaît ni la séquence ni la fonction mais dont on sait qu'il confère un phénotype important pour des sélectionneurs ou des agriculteurs, on peut utiliser la technique du clonage positionnel ou « marche chromosomique ». Cette technique permet de rechercher la localisation chromosomique d'un gène d'intérêt en s'appuyant sur des approches de génétique formelle. Elle peut se diviser en quatre parties (Figure 1.15) :

1) Cartographie génétique fine qui consiste à obtenir des marqueurs moléculaires qui entourent et qui soient les plus proches possibles du locus d'intérêt. Cette technique repose sur les événements de recombinaison génétiques qui se sont produits entre les marqueurs et le locus d'intérêt dans la population en ségrégation étudiée.

2) Cartographie physique qui consiste à identifier la région physique contenant le gène d'intérêt. Les marqueurs moléculaires permettent le positionnement de la carte physique sur la carte génétique. Pour les génomes non séquencés, la carte physique est matérialisée par des banques d'ADN de grands inserts clonés - YAC (Yeast Artificial Chromosomes), BAC (Bacterial Artificial Chromosomes), cosmides - qui collectivement contiennent la totalité de la région d'intérêt. La série de clones doit donc contenir des inserts chevauchants formant un ensemble complet de clones contigus que l'on appelle « contig ».

3) Identification de gènes candidats dans la région : cette étape passe par le séquençage et l'analyse du ou des BACs/YACs compris dans la fenêtre contenant le locus d'intérêt.

4) Validation du gène cloné : l'étape ultime du clonage positionnel est la démonstration que le gène identifié est bien celui qui confère le phénotype étudié. La démonstration la plus probante du clonage du gène recherché est la complémentation fonctionnelle par transformation génétique : dans le cas d'une résistance à un agent pathogène, la validation se traduira par la restauration de la résistance dans une variété sensible.

Le clonage positionnel est une stratégie longue et fastidieuse mais qui devient de plus en plus efficace chez les plantes modèles séquencées comme le riz, avec un génome diploïde de petite taille (Ueda *et al.*, 2005; Xu *et al.*, 2006; Saito *et al.*, 2010). Toutefois, cette technique reste un challenge majeur chez les espèces complexes à grand génome comme le blé tendre (16 000 Mb) ou la canne à sucre (10 000 Mb) à cause de deux contraintes principales : la taille du génome et la redondance génétique chez les plantes polyploïdes.

**Tableau 1.3. Exemples de banques BAC pour quelques espèces de Poacées**

Espèce		Taille génome (Mb)	Nombre de BACs	Taille moyenne des inserts (Kb)	Couverture du génome (X)	Référence
Riz (Nipponbare)		450	36 864	129	11	Chen <i>et al</i> , 2002
Sorgho (BTx623)		750	110 592	120	17	Begum, non publié
Maïs (B73)		2 500	247 680	136	13,5	Tomkins <i>et al</i> , 2002
Orge		5 000	313 344	106	6,3	Yu <i>et al</i> , 2000
Canne à sucre (R570)		10 000	103 296	130	1,3	Tomkins <i>et al</i> , 1999
Blé	<i>T.monococcum</i>	5 600	276 480	115	5,6	Lijavetzky <i>et al</i> , 1999
	<i>A.Tauschi</i>	4 400	114 000	119	3,7	Moulet <i>et al</i> , 1999
	<i>T.turgidum</i>	13 000	516 096	131	5,1	Cenci <i>et al</i> , 2003
	<i>T.aestivum</i>	16 000	1 200 000	130	9,3	Allouis <i>et al</i> , 2003



### D.1.1 La taille du génome

Dans la famille des Poacées, il existe une grande diversité de taille des génomes allant de 450 Mb pour le riz à 2 500 Mb pour le maïs, 5 000 Mb pour l'orge, 10 000 Mb pour la canne à sucre et 16 000 Mb pour le blé tendre. La variation de la taille de ces génomes peut s'expliquer par deux mécanismes :

(1) L'augmentation de taille des génomes peut être due à l'amplification/l'accumulation des éléments transposables (TE) (Bennetzen, 2005). Certains éléments transposables augmentent leur nombre de copies dans le génome lors de leur transposition par un mécanisme de « copier / coller » et ainsi augmentent la taille du génome. C'est le cas de la variété de riz diploïde *Oryza australiensis* (965 Mb) qui a une taille de génome deux fois supérieure au génome diploïde *Oryza sativa* (390 Mb). Cette variation de taille s'explique par la présence de trois familles de rétrotransposons qui représentent 605 Mb du génome d'*Oryza australiensis* (Piegu *et al.*, 2006).

(2) La duplication du génome entier, c'est-à-dire la polyploïdisation, provoque une augmentation majeure de la taille du génome (Doyle *et al.*, 2008).

Cette contrainte de taille des génomes intervient principalement lors de la construction des banques d'ADN à grands inserts. En effet, plus la taille du génome est grande, plus la banque devra contenir de clones pour avoir une couverture du génome suffisante pour trouver le caractère recherché (Tableau 1.3). Par exemple, la banque BAC de la variété *Sorghum bicolor* BTx623 (taille du génome de 750 Mb) contient 110 592 clones pour une couverture du génome de 17 équivalents génomes. En comparaison, la banque BAC du cultivar R570 (taille du génome 10 000 Mb) contient 103 296 clones pour une couverture du génome de 1,3 équivalent génome (Tomkins *et al.*, 1999).

### D.1.2 La redondance génétique

Par rapport au blé tendre, le génome de la canne à sucre est plus petit, mais son niveau de redondance est beaucoup plus élevé, avec douze haplotypes hom(é)ologues fortement hétérozygotes en moyenne à chaque locus par rapport à trois ensembles de paires d'haplotypes

**Tableau 1.4. Ratio de ségrégation d'un marqueur simplex, duplex, triplex dans une descendance issue d'une autofécondation, pour différents niveaux de polyploïdie (h) et d'appariements des chromosomes (Grivet, 1995).**

Type de comportement à la méiose	Marqueur simplex	Marqueur duplex	Marqueur triplex
Disomique (tous niveaux de polyploïdie)	3 : 1	15 : 1*	63 : 1*
Polysomique ( h=12 )	3 : 1	18,4 : 1	130 : 1
Polysomique ( h=10 )	3 : 1	19,2 : 1	163 : 1
Polysomique ( h=8 )	3 : 1	21 : 1	255 : 1
Polysomique ( h=6 )	3 : 1	24 : 1	1023 : 1
Polysomique ( h=4 )	3 : 1	35 : 1	-

\* Cas où les marqueurs duplex et triplex sont sur des chromosomes appartenant à des couples hom(é)ologues différents.

fortement homozygotes dans le blé tendre. Ce niveau très élevé de redondance génétique rend difficile le suivi des loci spécifiques en cartographie génétique et dans des approches de cartographie physique. Les conséquences de cette contrainte sont :

(1) L'utilisation de marqueurs simplex pour la cartographie génétique. Chez les diploïdes, chaque allèle d'un locus donné peut être présent en une seule copie si l'individu est hétérozygote ou en deux copies s'il est homozygote. Chez les polyploïdes d'un niveau de ploïdie  $h$ , le nombre de copies d'un allèle à un locus peut varier de 1 à  $h$ . On parlera donc d'un allèle simplex lorsqu'il est présent en une copie et d'un allèle multiplex lorsqu'il est présent en deux ou plusieurs copies (Wu *et al.*, 1992). Le ratio de ségrégation d'un marqueur varie en fonction du nombre de copies, du mode d'appariement des chromosomes à la méiose (comportement disomique ou polysomique) et du niveau de polyploïdie. Dans un comportement disomique ou polysomique, un marqueur dominant simplex ségrégera avec un ratio de 3 : 1 dans une descendance F2 (trois présences pour une absence). Le ratio de ségrégation d'un marqueur duplex dans un comportement disomique est 15 : 1 dans une descendance F2. Par contre, ce ratio pour le même marqueur dans un comportement polysomique est dépendant du niveau de polyploïdie de l'espèce (Tableau 1.4). Plus le niveau de polyploïdie est élevé et plus le ratio de ségrégation se rapprochera de 15 : 1. Pour les marqueurs multiplex, ce ratio augmente de façon importante. Pour un marqueur multiplex, la ségrégation dépend donc du type de comportement des chromosomes à la méiose et du niveau de polyploïdie. La visualisation de la ségrégation requiert un nombre de descendants d'autant plus grand que le rapport présence / absence est élevé. Par contre, les marqueurs simplex ne sont pas soumis à ces différents facteurs et sont donc plus facilement analysables.

(2) La différenciation des haplotypes hom(é)ologues sur la carte physique. Lors du criblage des banques d'ADN à grands inserts, les clones identifiés peuvent appartenir aux différents haplotypes constituant la région cible. Par exemple, pour une espèce dodécaploïde, 12 haplotypes hom(é)ologues peuvent recouvrir la région cible.

## **D.2. Stratégies déployées pour contourner ces contraintes**

:

### D.2.1 Utilisation d'un donneur diploïde ou d'un parent proche diploïde avec un plus petit génome

Cette stratégie a beaucoup été utilisée pour le clonage positionnel de gènes chez le blé tendre (Keller *et al.*, 2005). Elle consiste à utiliser des génomes de blé diploïdes ou tétraploïdes à plus petit génome qui sont étroitement liés au génome du blé tendre hexaploïde. Ainsi, pour le génome A du blé hexaploïde (16 000 Mb), on peut utiliser le génome A<sup>m</sup> du blé diploïde *T. monococcum* (5 600 Mb), alors que pour le génome D du blé hexaploïde, on peut utiliser l'espèce diploïde donneuse *Ae. tauschii* (4 400 Mb). Cette stratégie subgénomique proposée initialement par Gill *et al.*, (1991) a été une percée essentielle en matière de clonage positionnel sur le blé. Sur cette base plusieurs gènes ont été clonés comme :

- *Lr21* (Huang *et al.*, 2003), gène de résistance à la rouille brune du blé appartenant au génome D du blé tendre. La stratégie employée a été de réaliser un pont entre l'espèce diploïde *Ae. tauschii* et l'espèce hexaploïde *T. aestivum* pour réaliser la carte génétique. La carte physique a été réalisée à partir d'une banque de cosmides d'*Ae. tauschii*.
- *Lr10* (Feuillet *et al.*, 2003) est aussi un gène de résistance à la rouille brune. Deux gènes candidats ont été identifiés sur l'espèce diploïde *T. monococcum* et des sondes ont été dérivées de ces deux gènes pour cribler la banque BAC de l'espèce hexaploïde.
- *Pm3b* (Yahiaoui *et al.*, 2004) est un gène de résistance à l'oïdium. Ce gène a été cloné en dérivant des marqueurs de l'espèce diploïde *T. monococcum* et l'espèce tétraploïde *T. durum*. L'haplotype candidat dans l'espèce diploïde présentait une conservation partielle avec l'haplotype de l'espèce hexaploïde. Cette conservation a été suffisante pour dériver des marqueurs spécifiques et cribler la banque BAC de l'espèce hexaploïde.

Cette stratégie ne peut s'appliquer à toutes les plantes. Par exemple, chez la canne à sucre, nous ne connaissons pas d'espèces proches diploïdes dans le genre *Saccharum*. Seules des espèces polyploïdes sont connues.

### D.2.2 Synténie avec les plantes modèles

Cette stratégie est basée sur la synténie existante entre les génomes d'une même



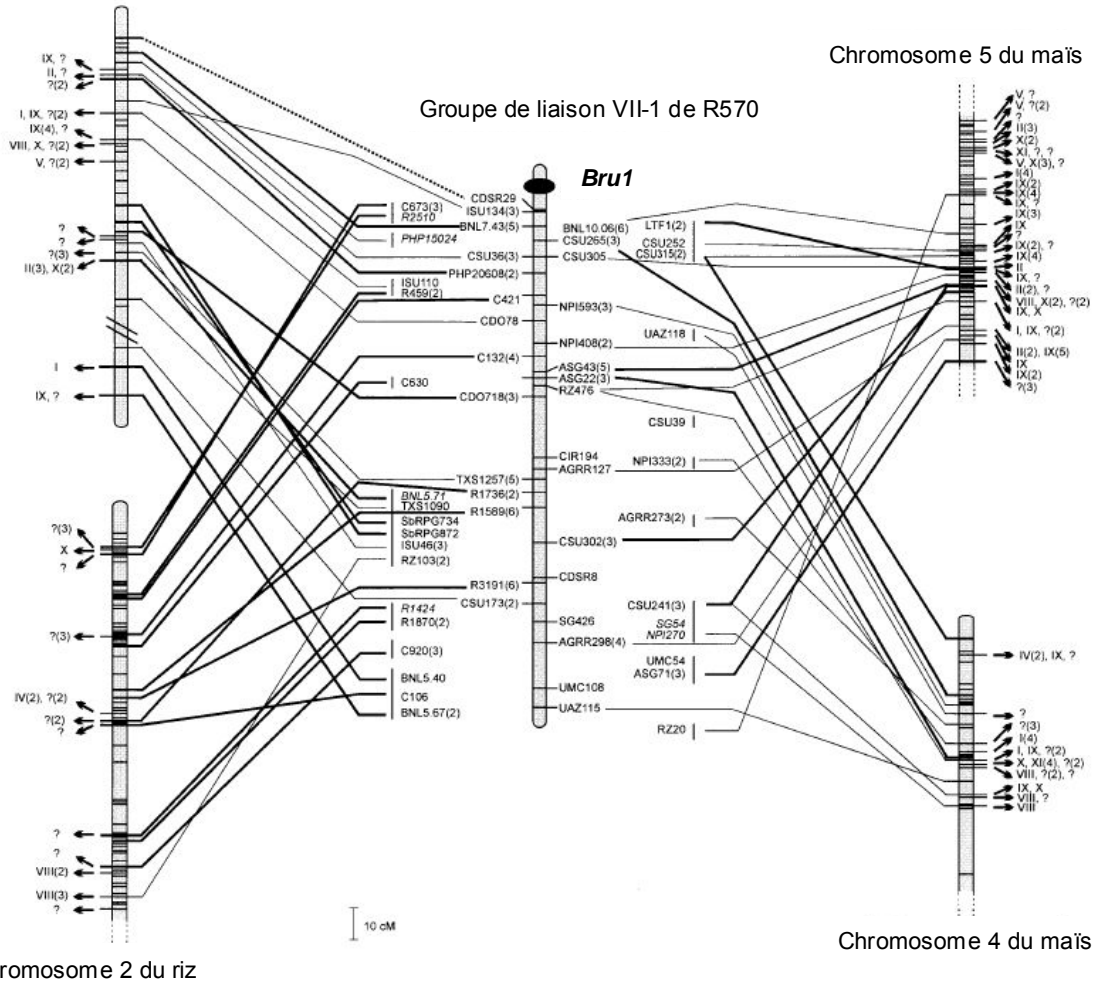
famille botanique, par exemple la famille des Poacées, la famille des Légumineuses ou la famille des Fabacées (Gale & Devos, 1998; Choi *et al.*, 2004; Panjabi *et al.*, 2008). Par comparaison des cartes génétiques avec une espèce modèle souvent diploïde et à petit génome, des marqueurs moléculaires vont être dérivés dans la région orthologue de l'espèce modèle pour densifier la carte génétique de la région du gène candidat dans l'espèce étudiée. Cette stratégie a été utilisée avec succès pour le clonage d'un gène de résistance chez l'orge : *Rpg1* (Kilian *et al.*, 1995; Han *et al.*, 1999; Brueggeman *et al.*, 2002). La synténie existante entre le riz et l'orge a été utilisée pour augmenter la densité de la carte génétique autour du gène *Rpg1*, même si *Rpg1* ne présente pas d'orthologue chez le riz (Brueggeman *et al.*, 2002). Cette stratégie a aussi été utilisée dans la construction de la carte fine du gène de résistance *Bru1* chez la canne à sucre (Le Cunff *et al.*, 2008). La synténie existante entre la canne à sucre et le sorgho, et, la canne à sucre et le riz, a été utilisée pour dériver des marqueurs moléculaires

### D.2.3 Analyse de Ségrégation en mélange

La technique BSA (Bulk Segregant Analysis) décrite par Michelmore (1991) est une méthode simple et rapide qui permet d'identifier des marqueurs moléculaires liés à un gène majeur. Elle consiste à comparer des mélanges (bulks) d'ADN constitués d'individus dont le phénotype est identique pour un caractère donné (par exemple résistant ou sensible pour une résistance à un pathogène) mais aléatoire pour les autres caractères. Cette technique permet donc d'obtenir des marqueurs locus-spécifiques. Elle est généralement associée à des techniques de marquage basé sur la PCR, comme les AFLP qui permettent de générer un grand nombre de marqueurs rapidement. La probabilité d'identifier des marqueurs liés à un gène avec cette méthode dépend de plusieurs facteurs, dont le taux de recombinaison entre le gène et le marqueur, la taille du bulk et le type de descendance.

Cette technique a été utilisée avec succès dans le cadre du clonage du gène de résistance à la rouille brune chez le blé, *Lr21* (Huang *et al.*, 2003). Plus récemment, un autre gène de résistance du blé à l'oïdium a été isolé grâce à cette technique (Xu *et al.*, 2010). Dans les premières étapes du clonage du gène *Bru1* de la canne à sucre, cette technique a permis

Groupe de liaison D du sorgho



**Figure 1.16. Relations synténiques entre le groupe de liaison VII du cultivar R570 et les régions orthologues chez le sorgho, le maïs et le riz (Asnaghi *et al.*, 2000).**

La position des sondes est indiquée par une ligne horizontale, courte lorsqu'elles ne s'hybrident pas sur l'ADN de canne à sucre, longue lorsqu'elle révèle un marqueur sur le groupe d'homologie VII et/ou un autre groupe d'homologie. Les lignes horizontales ou diagonales entre les fragments chromosomiques indiquent des loci colinéaires entre la canne et les autres graminées. Les lignes épaisses indiquent les sondes qui sont précisément localisées sur les segments orthologues et les lignes fines indiquent celles dont la localisation est plus imprécise.



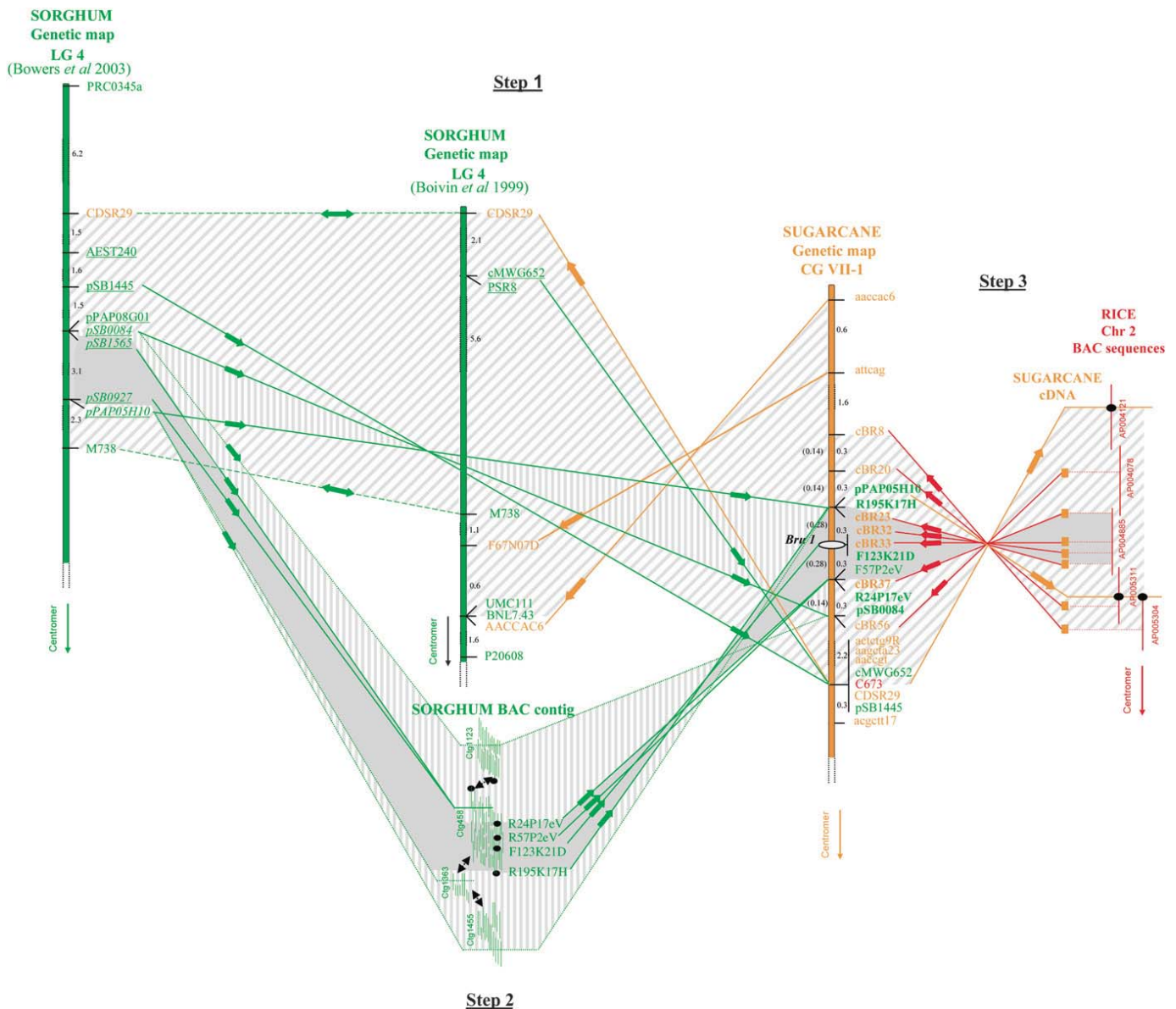
d'identifier des marqueurs entourant le gène (Asnaghi *et al.*, 2004).

### **E. Etapes et stratégie mise en place pour le clonage positionnel de *Bru1***

Un gène majeur de résistance à la rouille brune de la canne à sucre, *Bru1*, a été identifié chez le cultivar réunionnais R570 (Daugrois *et al.*, 1996). La première cartographie génétique de ce cultivar, réalisée avec une population de 77 individus issus d'une autofécondation, a montré que ce gène était lié avec un marqueur RFLP, CDSR29, indépendant de la carte (Grivet *et al.*, 1996). Une étude du spectre d'action de ce gène a démontré que ce gène confère une résistance à la rouille brune pour huit isolats provenant de régions géographiques différentes (Brésil, Colombie, Floride, Guadeloupe, Réunion et Zimbabwe) (Asnaghi *et al.*, 2001). De plus, le cultivar R570 est résistant à la rouille dans l'ensemble des zones où il est cultivé (La Réunion, Maurice, Guadeloupe, en Afrique de l'Ouest et de l'Est. Le gène *Bru1* semble donc conférer une résistance durable à la rouille brune de la canne à sucre. Ces éléments ont permis d'entreprendre un projet de clonage positionnel, ce qui représente un véritable challenge étant donné la complexité du génome de la canne à sucre.

La première approche a consisté à exploiter les relations synténiques entre le sorgho, le riz, le maïs et la canne à sucre, en utilisant la population de cartographie de Grivet *et al.* (1996). La sonde CDSR29 a été cartographiée chez le sorgho sur le groupe de coségrégation D défini par Boivin *et al.* (1999). Les régions orthologues chez le riz et chez le maïs ont été identifiées. Elles correspondent à l'extrémité du chromosome 2 du riz et à l'extrémité du chromosome 4 et à la partie centromérique du chromosome 5 du maïs. Les sondes dans ces régions ont permis de définir le groupe de coségrégation VII-1 sur lequel se situe le gène *Bru1* (Asnaghi *et al.*, 2000) (Figure 1.16). Cependant, le marqueur moléculaire le plus proche (CDSR29) restait à 2,2 cM du gène *Bru1*.

Pour affiner la carte génétique, une population de 312 individus issus de l'autofécondation a été développée. Dans un premier temps, la méthode de BSA couplée à la



**Figure 1.17. Exploitation de la synténie avec le sorgho et le riz pour saturer la carte génétique de la région de *Bru1* (Le Cunff *et al.*, 2008)**

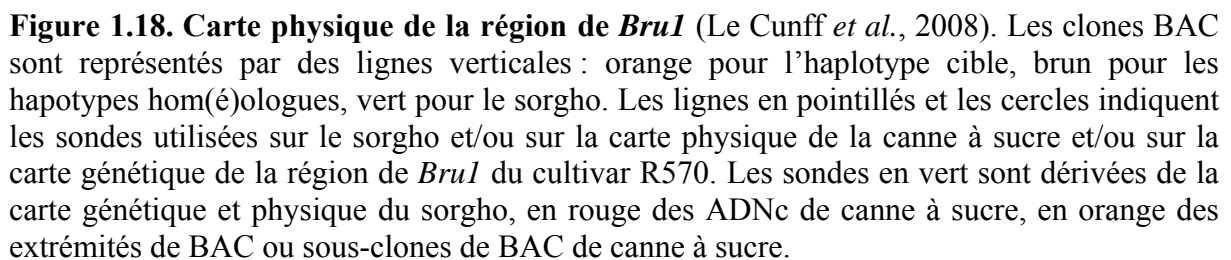
(Step 1) Exploitation des cartes génétiques du sorgho. Les marqueurs RFLP en orange correspondent à des marqueurs de canne à sucre. Les doubles flèches vertes indiquent les marqueurs communs entre les deux cartes de sorgho. Les flèches vertes indiquent les marqueurs RFLP dérivés du sorgho (en vert) cartographiés sur la descendance de R570. (Step2) Exploitation d'une carte physique partielle du sorgho. Les doubles flèches noires indiquent les liens entre les contigs. (Step3) Exploitation de la séquence de 5 BACs de riz (lignes verticales rouges) et d'une banque d'ADNc de canne à sucre pour définir de nouveaux marqueurs (lignes et flèche rouges).

Les distances sont indiquées en centiMorgan. Pour la carte de la canne à sucre, les distances indiquées sur la droite sont basées sur 312 individus et les distances entre parenthèses sur la gauche sont basées sur 712 individus.

technique AFLP a été utilisée pour augmenter la densité en marqueurs dans la région cible. Cette approche a permis de cartographier huit marqueurs AFLP dont quatre de chaque côté du gène *Bru1*. Les marqueurs les plus proches sont cartographiés à 1,9 cM en position distale et 2,2 cM en position proximale par rapport au gène *Bru1* (Asnaghi *et al.*, 2004). Dans un second temps, deux des marqueurs AFLP générés par la méthode BSA situés en position distale ont été clonés et cartographiés sur la carte du sorgho de Boivin *et al.*, (1999) pour mieux délimiter la région du sorgho orthologue à la région cible chez la canne à sucre. Un des deux marqueurs s'est cartographié en position proximale par rapport au marqueur CDSR29 sur le groupe de liaison D. L'autre marqueur a présenté un profil monomorphe et a été utilisé pour cribler une banque BAC du sorgho construite à l'université de Clemson. Cette sonde a permis d'identifier entre autre le BAC 67N07 dont une de ses extrémités terminales, F67N07, se cartographie aussi en position proximale par rapport à CDSR29 sur le groupe de liaison D du sorgho. La région du sorgho orthologue à la région cible de la canne à sucre a pu être délimitée sur le groupe de liaison D et est inversée entre le génome du sorgho et le génome du cultivar moderne R570.

La cartographie génétique fine du cultivar R570 a ensuite été poursuivie en exploitant les relations synténiques entre le sorgho, le riz et la canne à sucre (Le Cunff *et al.*, 2008). Chez le sorgho, deux marqueurs de la carte de Boivin *et al.*, (1999) sont communs avec l'extrémité du groupe de liaison F de la carte génétique de sorgho de Bowers *et al.*, (2003). Ces deux sondes ont permis de délimiter la région orthologue de la région cible sur la carte de Bowers *et al.*, (2003). Cinq nouveaux marqueurs ont ainsi pu être cartographiés dans la région cible de R570. Ces marqueurs ont permis le criblage d'une banque BAC sorgho partiellement ordonnée et la construction d'un contig couvrant la région orthologue à la région cible. Les extrémités des BACs du contig sorgho ont représenté une nouvelle source de marqueurs. L'ensemble de ce travail a permis de cartographier six marqueurs autour de *Bru1* dont trois situés à 0,3 cM de chaque côté du gène d'intérêt et un marqueur qui coségrège avec le gène (Figure 1.17) (Le Cunff *et al.*, 2008).

La région du riz (extrémité du chromosome 2) orthologue à la région cible chez la canne à sucre a été définie en criblant une banque BAC de riz avec certains des marqueurs encadrant le gène *Bru1*. Cinq clones BAC de riz couvrent la région cible (Figure 1.17). La séquence de ces cinq BAC est comparée à la banque de 250 000 EST de canne à sucre développée au Brésil (Vettore *et al.*, 2003). Les ADNc, correspondant aux EST, présentant un fort taux d'homologie avec la séquence de ces BAC riz ont été utilisés comme sondes pour



affiner la carte génétique. Sept nouveaux marqueurs ont ainsi été cartographiés, dont quatre se situent à 0,3 cM de part et d'autre de *Bru1* et deux coségrègent avec le gène (Figure 1.18) (Le Cunff *et al.*, 2008).

Pour augmenter la résolution de la carte génétique, de nouveaux individus recombinants ont été recherchés dans une nouvelle population issue de l'autofécondation de R570. La moitié des individus recombinants étant sensibles (autofécondation) et leur phénotype étant plus facilement identifiable, seuls ces individus (400 sur 1600) ont été génotypés en utilisant deux marqueurs AFLP, situés de chaque côté du gène. Les individus sélectionnés comme présentant une recombinaison dans la région cible ont ensuite été analysés avec les marqueurs RFLP situés dans une fenêtre, plus étroite, autour de *Bru1*. Les distances génétiques ont alors été recalculées en fonction des nouvelles recombinaisons trouvées dans la région cible. Finalement, trois marqueurs sont situés en position distale et trois en position proximale par rapport au gène, chacun à une distance de 0,28 cM de *Bru1*, et trois marqueurs coségrègent avec le gène (Figure 1.18).

A cette étape du projet la construction d'une carte physique de la région cible a été initiée par le criblage d'une banque BAC de canne à sucre. Cette étape, pour une espèce diploïde homozygote, doit aboutir à l'identification d'un unique contig. Chez la canne à sucre, espèce polyploïde hétérozygote, les BAC identifiés peuvent appartenir à autant de contigs que de chromosomes hom(é)ologues. Une étape supplémentaire est donc nécessaire pour identifier les BAC qui appartiennent à l'haplotype cible porteur du gène *Bru1*. Le criblage de la banque BAC du cultivar R570 (réalisée à l'Université de Clemson) (Tomkins *et al.*, 1999) a été réalisé avec neuf marqueurs cartographiés à proximité de *Bru1*. Les BACs identifiés appartenaient à sept haplotypes différents. Un seul BAC couvrait complètement la région cible pour certains haplotypes hom(é)ologues. En revanche, seul un BAC appartenait à l'haplotype porteur de *Bru1* et ne couvrait pas la totalité de la région cible. Une nouvelle banque a donc été construite au CIRAD à partir de quatre descendants issus de l'autofécondation du cultivar R570 et présentant chacun deux copies du gène *Bru1* (Le Cunff *et al.*, 2008). Trois nouveaux BAC de l'haplotype cible ont été identifiés mais ils ne se chevauchaient pas. Il reste donc deux espaces au niveau de la région cible, non couverts par les BAC. Une marche chromosomique a donc été initiée en vue de couvrir complètement la région contenant *Bru1*. Celle-ci consiste à utiliser les extrémités des BAC ou des sous-clones de BAC de la région cible pour identifier par criblage de nouveaux BAC chevauchants.



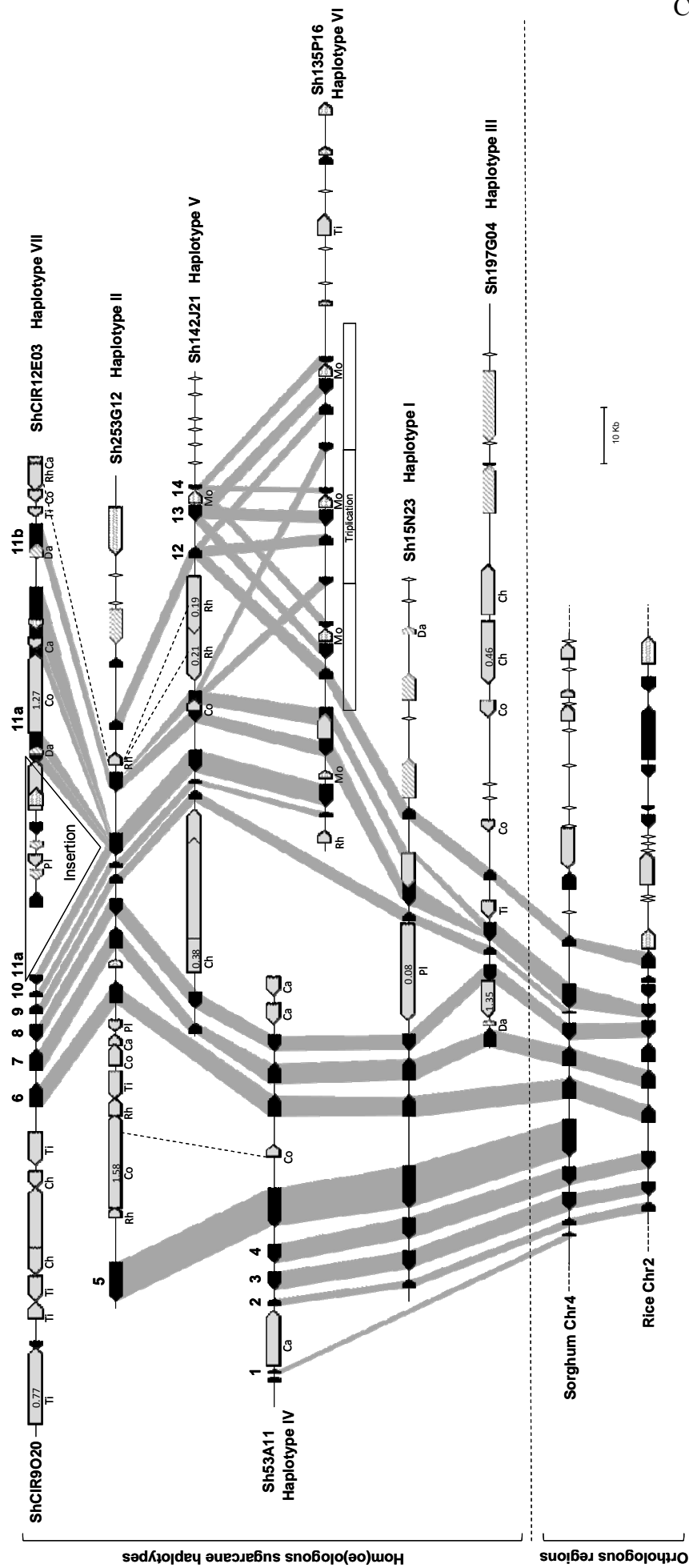
Au final, huit BAC formant trois contigs appartenant à l'haplotype cible ont été identifiés mais deux espaces non couverts demeurent (Figure 1.19). Parallèlement à la construction de la carte physique, la carte génétique a été affinée dans la région cible avec 13 nouveaux marqueurs (sous-clones BAC) dont 12 coségrègent avec *Bru1*. En effet, chaque BAC nouvellement identifié et appartenant à la région cible est une nouvelle source potentielle de marqueurs pour la cartographie génétique (Figure 1.18) (Le Cunff *et al.*, 2008). D'autre part, certaines sondes RFLP développées à partir des BAC de l'haplotype cible ne s'hybrident pas sur les BAC appartenant aux haplotypes hom(é)ologues et présentent un profil particulier avec une seule bande contrairement aux multiples bandes généralement observées chez ce haut polyploïde. Ceci montre la présence d'une insertion dans l'haplotype cible. L'utilisation de 2000 individus supplémentaires issus de l'autofécondation de R570 pour réduire la région de *Bru1* n'a pas permis de trouver de nouveaux recombinants (données non publiées). En effet, cette insertion entraîne probablement une diminution du taux de recombinaison dans la région cible (Le Cunff *et al.*, 2008).

La carte physique de la région comprend 15 clones BAC regroupés en sept haplotypes hom(é)ologues dont un correspond à l'haplotype cible porteur du gène *Bru1* (Figure 1.19). Un ensemble de 13 clones BAC dont quatre correspondent à l'haplotype cible, six aux haplotypes hom(é)ologues et trois à la région orthologue chez le sorgho ont été séquencés et annotés (Centre National de Séquençage, Evry, France).

## **F. Organisation du locus de *Bru1***

Cette partie est le résumé d'un article accepté dans la revue *New Phytologist* intitulé « High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane » (Garsmeur *et al.*, 2010) (Annexe 1).

Le séquençage et l'annotation de huit BACs correspondant aux sept haplotypes hom(é)ologues de la région de *Bru1* a permis de réaliser une comparaison de l'organisation physique des différents haplotypes hom(é)ologues entre eux mais aussi avec la région orthologue du sorgho et du riz (Garsmeur *et al.*, 2010). Les 960 kb de séquence totale sont constitués de 337 kb d'éléments répétés, 207 kb de gènes et 416 kb de région intergénique



**Figure 1.19. Comparaison de l'organisation physique entre sept clones BAC hom(é)ologues de canna à sucre et avec les régions orthologues du sorgho et du riz da la région de Bru1 (Garsmeur *et al.*, 2010).**

Les gènes sont indiqués par des boîtes noires, LTR-rétrotransposons par des boîtes grises (le temps d'insertion est indiqué à l'intérieur), non-LTR rétrotransposons par des boîtes grises striées, transposons par des boîtes grises en pointillées et les « protéinase inhibitor » par des losanges blancs. Les gènes colinéaires sont reliés entre eux par des surfaces grises et les éléments transposables colinéaire par des lignes en pointillées. L'abréviation des familles des éléments transposables correspond à Rh (Rhum), Ti (TiPunch), Da (Daiquiri), Ca (Cachirinha), Ch (Cachaca), Co (Colada), PI (Planteur), Mo (Mojito).



(Garsmeur *et al.*, 2010). L'annotation des séquences a révélé la présence de 50 gènes et 66 éléments transposables répartis sur les différents haplotypes hom(é)ologues (Figure 1.19). Il s'agit donc d'une région riche en gènes, avec en moyenne un gène tous les 9 kb. La fonction de 15 gènes a pu être attribuée sur la base d'alignements protéiques avec les séquences disponibles dans les bases de données. Tous ces gènes, excepté le gène 10, sont prédits *in silico* comme étant fonctionnels.

Il a été montré l'existence d'un niveau très élevé de colinéarité des gènes entre les différents haplotypes hom(é)ologues de canne à sucre et une très bonne conservation de la structure et de la séquence des gènes : 78,8 à 100 % d'identité nucléotidique des séquences codantes, avec une moyenne de 95,9 % (Garsmeur *et al.*, 2010) (Figure 1.19). Néanmoins, certains fragments entraînent une perturbation de cette colinéarité : (i) la présence de gènes supplémentaires entre les gènes 11a et 11b sur l'haplotype VII, correspondant à l'insertion du fragment d'ADN (Le Cunff *et al.*, 2008) (ii) la présence d'un fragment de 22 kb tripliqué sur l'haplotype VI incluant une partie du gène 11b et les gènes 12, 13 et 14.

Un niveau élevé de synténie avec la région orthologue du sorgho a également été observé (91,6 % d'identité nucléotidique en moyenne pour les séquences codantes), malgré l'absence des gènes 9, 10, 11a, 13 et 14. Une colinéarité plus faible avec la région orthologue du riz a été observée (71 % d'identité nucléotidique en moyenne pour les séquences codantes). Les gènes 1, 5, 9, 10, 11a, 13 et 14 sont absents dans cette région chez le riz.

Les LTR rétrotransposons représentent le groupe majoritaire des éléments transposables dans la région de *Bru1* (65 % de tous les TE). Vingt pour cent des TE sont complets (12 LTR rétrotransposons et 1 transposon). La majorité des TE sont localisés dans des domaines intergéniques, excepté pour huit d'entre eux qui sont insérés dans un intron d'un allèle du gène 11a et de quatre allèles du gène 11b (Figure 1.19). Contrairement aux gènes de cette région, on observe très peu de colinéarité entre les TE des différents haplotypes hom(é)ologues. Aucun TE n'est présent entre les gènes 1 et 12 des régions orthologues du sorgho et du riz (Garsmeur *et al.*, 2010).

Ces données acquises sur la région de *Bru1* confirment les résultats d'une étude précédente qui comparait deux haplotypes hom(é)ologues correspondant à une région portant un gène codant une alcool deshydrogénase Adh 1 chez la canne à sucre (Jannoo *et al.*, 2007).



La canne à sucre ne semble donc pas avoir subi de remaniement majeur de son génome en dépit de son haut niveau de polyploïdie.

## **G. Objectifs de la thèse**

Les recherches menées dans l'équipe de l'unité Développement et Amélioration des Plantes du CIRAD de Montpellier où j'ai effectué ce travail de thèse visent à mieux comprendre la structure et l'évolution des génomes chez les monocotylédones *via* une approche de génomique comparative entre des espèces complexes telle que la canne à sucre et des espèces modèles comme le riz et le sorgho. L'objectif appliqué est de contribuer à l'acquisition de connaissances et d'outils pour l'amélioration des espèces cultivées tropicales. L'équipe s'intéresse en particulier à la caractérisation d'un gène majeur de résistance à la rouille (*Bru1*) chez le cultivar moderne de canne à sucre R570, par une approche de clonage positionnel et en exploitant la bonne relation de synténie entre les génomes de la canne à sucre et des graminées modèles sorgho et riz. Cette approche a permis à mon équipe de construire une carte génétique fine et la première carte physique de la région de *Bru1*, qui comprend sept haplotypes hom(é)ologues dont un correspondant à l'haplotype cible porteur du gène *Bru1* (Le Cunff *et al.*, 2008). Cet haplotype est composé de sept clones BAC qui ne se chevauchent que partiellement, laissant deux espaces non couverts. Cette situation résulte de la présence d'une insertion dans la séquence de l'haplotype porteur de *Bru1*.

La canne à sucre possède l'un des génomes le plus complexe étudié à ce jour chez les espèces cultivées, en particulier de part son haut niveau de polyploïdie ( $2n \approx 12x \approx 115$ ), son aneuploïdie et son origine interspécifique. La stratégie du clonage positionnel représente donc un véritable défi chez cette espèce, avec deux contraintes principales : la taille du génome et le niveau de redondance génétique. A ce jour, aucun gène majeur de résistance n'a été cloné chez la canne à sucre.

Dans ce contexte, ce projet de thèse s'est centré sur la structure de ce locus de résistance à la rouille et plus particulièrement sur la cartographie physique de l'haplotype cible porteur du gène *Bru1*. Cette thèse s'articule autour de trois objectifs qui formeront les trois chapitres de résultats :



- Poursuivre la cartographie physique de l'haplotype cible en exploitant les relations synténiques entre les différents haplotypes hom(é)ologues de la région de *Bru1* afin de combler les deux espaces non encore couverts.
- Identifier et caractériser les gènes candidats potentiels présents sur l'haplotype cible dans la région de *Bru1* grâce au séquençage des clones BAC.
- Déterminer l'origine du fragment inséré dans l'haplotype cible de la région de *Bru1*. L'étude de la structuration de l'insertion dans diverses espèces de canne à sucre nous permettra de répondre à plusieurs questions : quelle(s) est (sont) le(s) espèce(s) du genre *Saccharum* spp. qui possèdent l'insertion ? Quelle(s) espèce(s) a (ont) transmis cette insertion aux cultivars modernes et à quel moment au cours de l'amélioration variétale ?

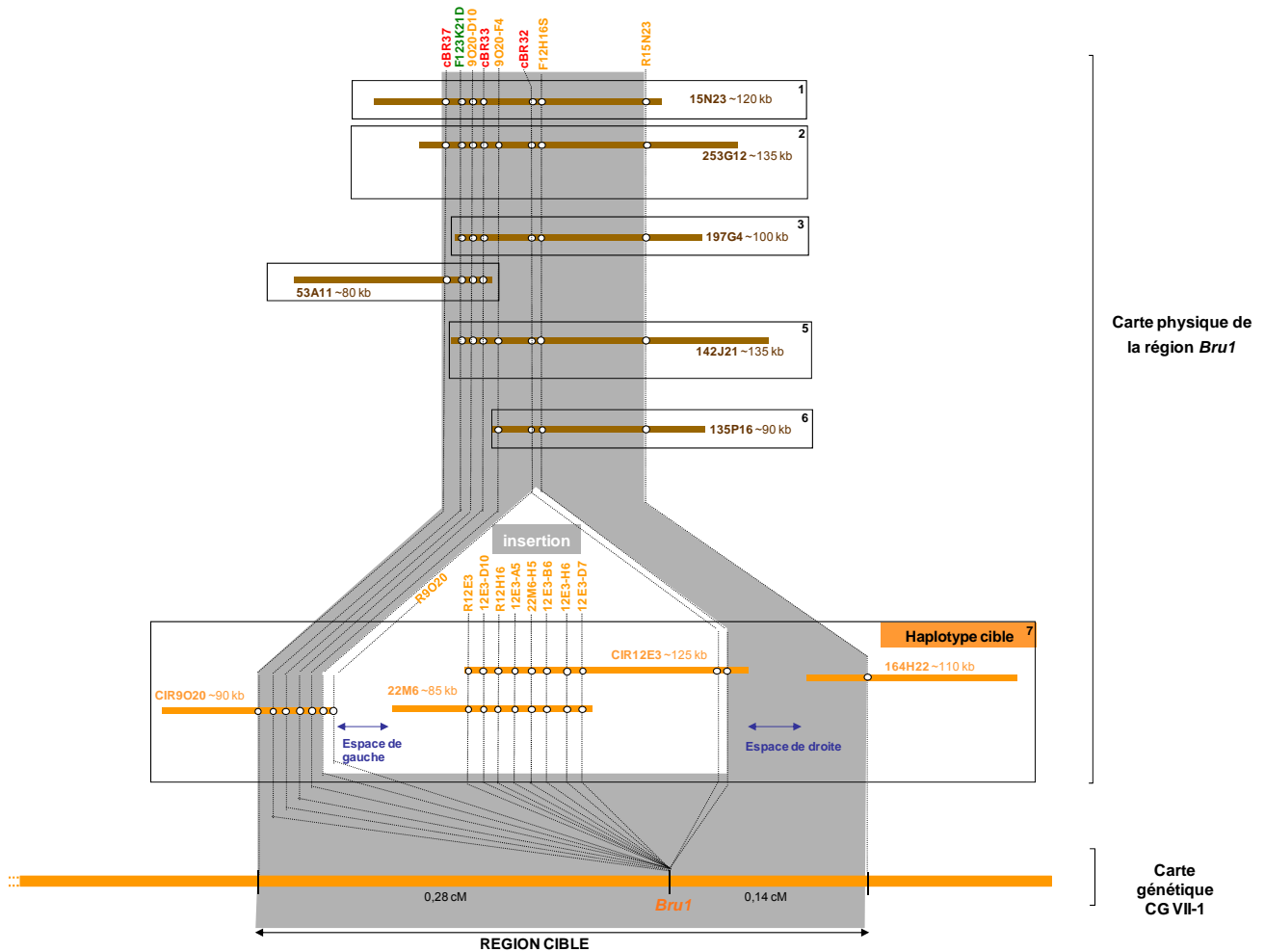
Une discussion générale portant sur l'ensemble de mes résultats et les perspectives de ce travail constituent le cinquième chapitre.



---

## **Chapitre 2. Cartographie physique de l'haploype cible**

---



**Figure 2.1. Carte physique de la région de *Bru1*** (Le Cunff *et al.*, 2008). Les clones BAC sont représentés par des lignes verticales : orange pour l'haplotype cible, brun pour les haplotypes hom(é)ologues, vert pour le sorgho. Les lignes en pointillées et les cercles indiquent les sondes utilisées sur le sorgho et/ou sur la carte physique de la canne à sucre et/ou sur la carte génétique de la région de *Bru1* du cultivar R570. Les sondes en vert sont dérivées de la carte génétique et physique du sorgho, en rouge des ADNc de canne à sucre, en orange, des extrémités de BAC ou sous-clones de BAC de canne à sucre.



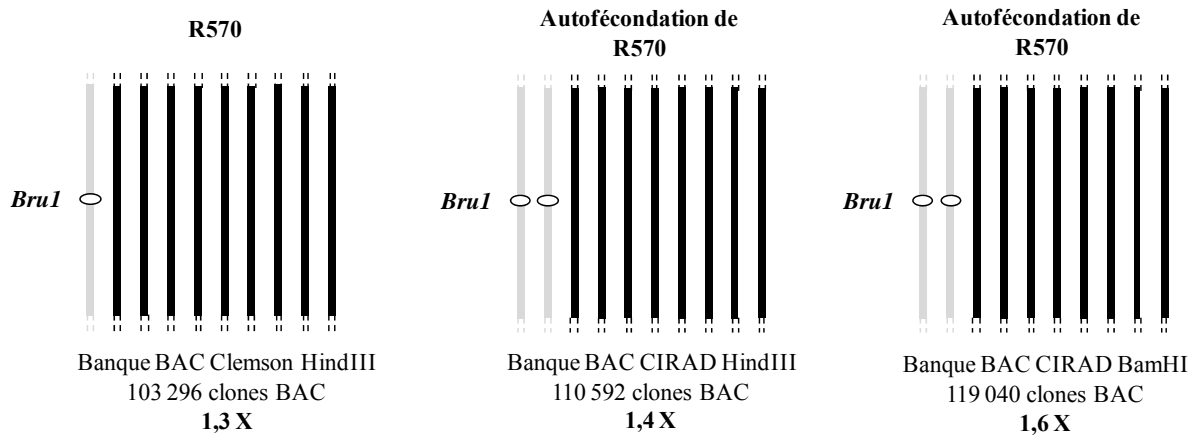
## Chapitre 2 : Cartographie physique de l'haplotype contenant le gène *Bru1*.

### A. Stratégies

#### A.1. Stratégies générales

La carte physique de la région de *Bru1* disponible au début de ma thèse comprenait 15 clones BAC regroupés en sept haplotypes hom(é)ologues dont un correspondant à l'haplotype cible porteur du gène *Bru1* (Figure 2.1). Pour tous les haplotypes, un seul clone BAC comprend l'ensemble de la région cible (entre les marqueurs cBR37 et R15N23), excepté pour l'haplotype cible qui comprend quatre clones BAC qui ne se chevauchent que partiellement, laissant deux espaces non couverts nommés « espace de gauche » et « espace de droite » (Figure 2.1). Cette situation résulte notamment de la présence d'une insertion dans l'haplotype porteur de *Bru1* (Le Cunff *et al.*, 2008). Le séquençage et l'annotation de l'ensemble de ces BAC ont montré qu'il existe une importante conservation de la synténie et de la colinéarité au niveau des gènes et des régions intergéniques mais pas au niveau des éléments transposables (Garsmeur *et al.*, 2010). D'autre part, ces données ont permis de démontrer que l'espace de gauche est inclus dans l'insertion spécifique de l'haplotype cible. Sur la base de la colinéarité entre les haplotypes, nous avons estimé la taille de l'espace de droite à environ 30 kb. La taille de l'espace de gauche ne peut en revanche pas être estimée du fait de la présence d'une insertion. Pour combler l'espace de droite, la synténie existant avec les autres haplotypes de la région de *Bru1* a été exploitée. Par contre, cette stratégie n'était pas envisageable pour combler l'espace de gauche et nous avons cherché une région orthologue chez le sorgho. Notre stratégie générale pour combler ces deux espaces non couverts a donc été :

- 1) d'exploiter les données issues du séquençage et de l'annotation des BAC de la région de *Bru1* pour définir de nouvelles sondes dans les régions flanquant les deux espaces non couverts de la carte physique et dans la région homologue sur les autres haplotypes pour l'espace de droite afin de cribler les deux banques BAC HindIII existantes.



---

**Recouvrement de la région cible**  
 $= 1,3 X + 2(1,4 X) + 2(1,6 X) = 7,3 X$

**Figure2.2. Représentation de l'haplotype cible dans les 3 banques BAC.**

2) de construire et cribler une nouvelle banque BAC pour augmenter la couverture de la région de *Bru1*.

### A.2. Ressources BAC utilisées et développées

Deux banques BAC réalisées avec l'enzyme HindIII étaient disponibles au début de ma thèse. Nous en avons construit une troisième avec l'enzyme BamHI. Les banques utilisées (Figure 2.2) sont :

- Une banque BAC du cultivar R570 construite à l'université de Clemson (Tomkins *et al.*, 1999) contenant 103 296 clones avec une taille moyenne des inserts de 130 kb, soit un taux de recouvrement de 1,2 fois le génome total. L'ensemble de la banque est contenu sur six filtres haute densité.

- Une banque BAC construite au CIRAD avec quatre individus résistants à la rouille issus de l'autofécondation du cultivar R570 possédant deux doses de l'haplotype cible (Le Cunff *et al.*, 2008). Ces individus ont été utilisés afin de multiplier par deux la représentation de l'haplotype cible dans la banque. En effet, le cultivar R570 possédant une copie du gène *Bru1*, la descendance issue de son autofécondation se compose de 25% d'individus ne possédant pas le gène *Bru1*, de 50% possédant une copie du gène *Bru1* et de 25% possédant deux copies du gène *Bru1*. Ces quatre individus ont été identifiés en comparant leur profil RFLP (intensité deux fois supérieure des bandes) avec celui de R570 pour quatre marqueurs encadrant le gène *Bru1*. Les clones BAC ont été regroupés par six. Cette banque comprend 110 592 clones avec une taille moyenne des inserts de 130 kb, soit un taux de recouvrement de 1,4 fois le génome total et de 2,8 fois la région de *Bru1*. Les BAC provenant de cette banque sont nommés CIR + coordonnées du BAC. L'ensemble de la banque est contenu sur un filtre haute densité.

- Une banque BAC que nous avons construite avec l'enzyme de restriction BamHI, selon le protocole de Vilarinhos *et al.*, (2003). Cette banque a été réalisée avec les quatre mêmes individus issus de l'autofécondation de R570 porteurs d'une double dose de l'haplotype cible. Elle contient 119 040 clones avec une taille moyenne des inserts de 130 kb.



Le taux de recouvrement est de 1,6 fois le génome total et de 3,2 fois la région de *Bru1*. Les BAC provenant de cette banque sont nommés CIRB + coordonnées du BAC. L'ensemble de la banque est contenu sur six filtres haute densité.

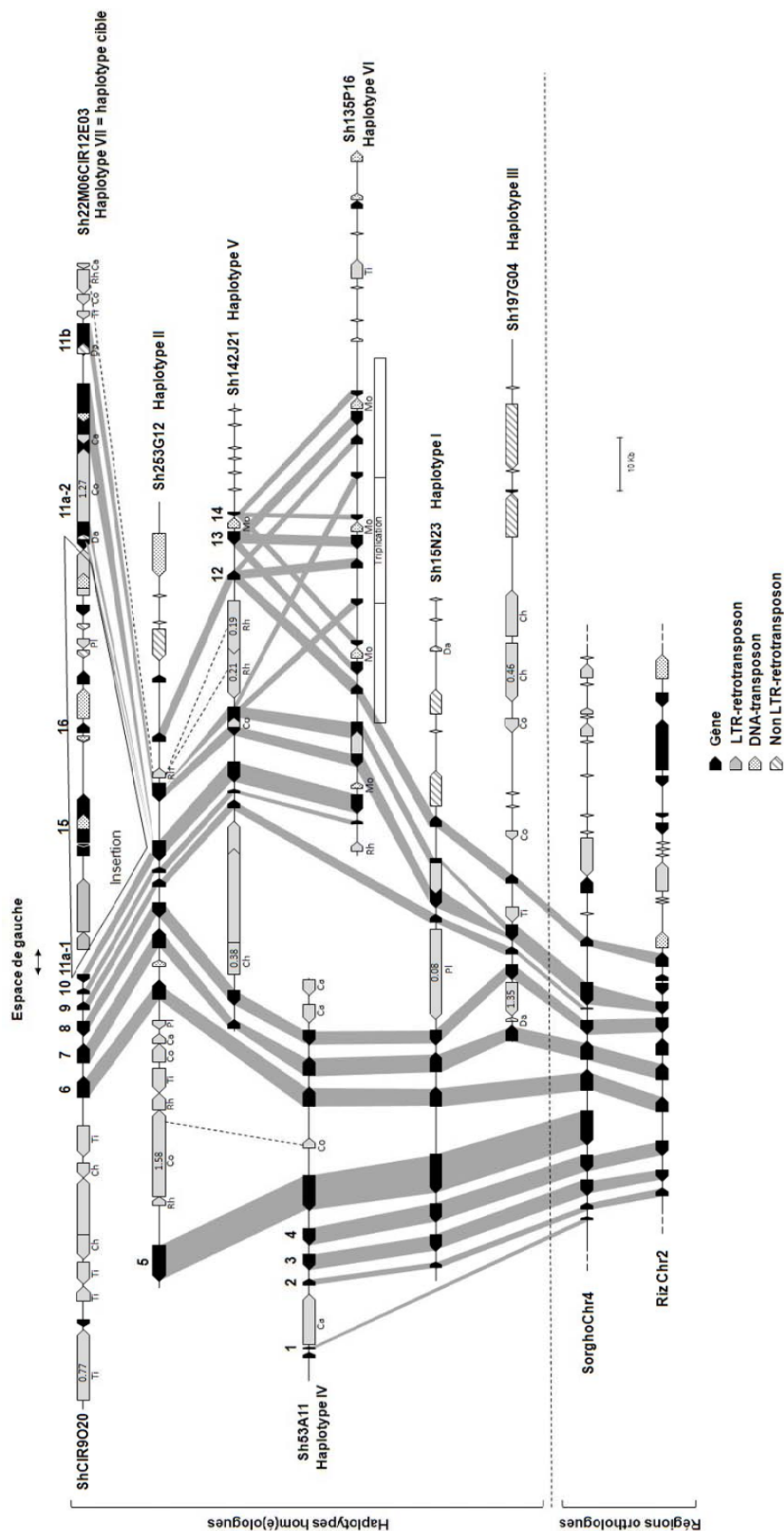
Nous disposons donc de 3 banques BAC représentant une couverture totale du génome du cultivar R570 de 4,2 X et une couverture de la région de *Bru1* de 7,3 X (Figure 2.2).

### **A.3. Principe d'identification des clones BAC correspondant à l'haplotype cible**

Pour cribler les banques BAC, nous avons hybridé les filtres haute densité avec des sondes définies dans les régions flanquantes des deux espaces non couverts de la carte physique de l'haplotype cible, sur la base de la conservation de la synténie entre les différents haplotypes hom(é)ologues de la région. La polyploïdie et l'hétérozygotie de la canne à sucre sont deux contraintes majeures pour la construction d'une carte physique. En effet, les BAC identifiés par hybridation appartiennent aux différents haplotypes hom(é)ologues couvrant la région d'intérêt et donc potentiellement à autant de contigs de BAC qu'il y a d'haplotypes hom(é)ologues. En comparaison, chez une plante diploïde homozygote, les BAC identifiés correspondraient à un unique contig de BAC. Donc il est nécessaire de définir une stratégie pour cribler nos ressources BAC et tester l'appartenance d'un BAC à l'haplotype cible. Notre méthode est basée sur trois points :

1) Vérification par amplification PCR : les couples d'amorces ayant servi à faire les sondes et les autres marqueurs PCR disponibles dans la région de *Bru1* sont utilisés pour l'étape de vérification par amplification PCR des BAC hybridés par les sondes. Certains de ces marqueurs PCR sont spécifiques de l'haplotype cible et dans la suite du texte seront identifiés par leur nom suivi d'un astérisque (\*).

2) DNA Fingerprinting : un profil de digestion du clone BAC est réalisé et comparé aux profils des BAC de l'haplotype cible déjà connus. Ceci permet de vérifier la présence d'un chevauchement entre le BAC candidat et un des BAC appartenant à l'haplotype cible. Le fingerprint nous permet aussi de déterminer approximativement la taille du fragment couvrant un des deux espaces. Toutefois, l'interprétation des fingerprint n'est pas toujours facile quand nous devons comparer des BAC clonés avec des enzymes différentes (HindIII, BamHI).



**Figure 2.3. Organisation physique des 7 clones BAC hom(é)ologues au niveau de la région de *Bru1* et des segments orthologues chez le sorgho et le riz (d'après Garsmeur et al., 2010).** Les gènes sont indiqués par des boîtes noires, les rétrotransposons LTR par des boîtes grises (temps d'insertion à l'intérieur des boîtes), les rétrotransposons non-LTR par des boîtes hachurées, les transposons par des boîtes pointillées. Les traits pleins grisés représentent les relations synténiques entre les gènes (numérotés de 1 à 14) des différents haplotypes. Les traits en pointillés représentent les relations synténiques entre les éléments transposables.

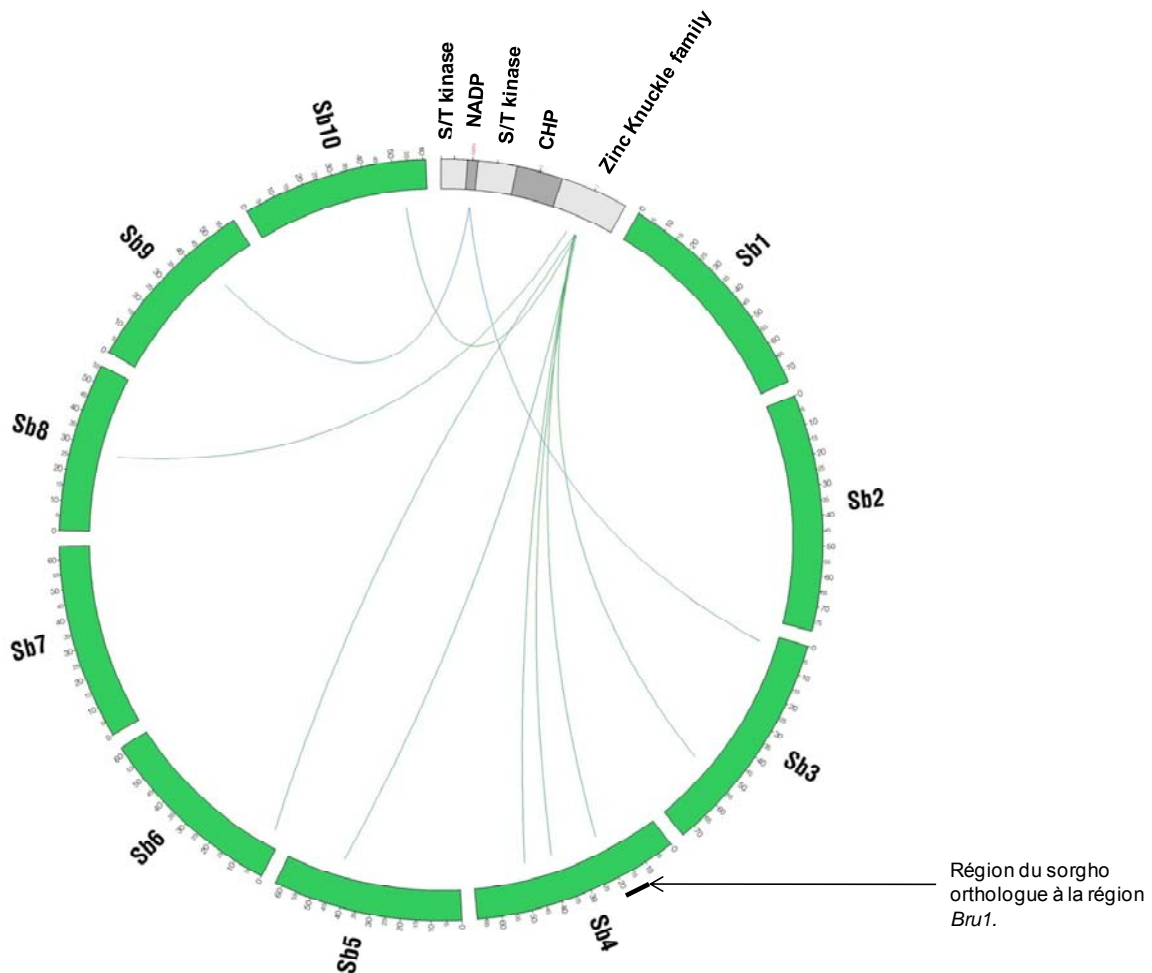
3) Séquençage des extrémités des BAC (BES) : les deux extrémités des BAC candidats sont séquencées. Les séquences sont ensuite comparées avec les séquences connues des BAC hom(é)ologues de la région de *Bru1* pour rechercher un point d'ancrage et déterminer si ce BAC correspond à un des haplotypes connus de la région de *Bru1*. Ces séquences sont aussi comparées avec les BES déjà séquencées et avec la séquence des génomes complets du sorgho et du riz, afin d'identifier la région orthologue potentielle chez ces autres espèces.

## **B. Cartographie physique pour compléter l'espace de gauche**

La carte physique de l'haplotype cible décrite sur la figure 2.3 montre que l'espace de gauche se situe dans l'insertion présente sur l'haplotype cible (Le Cunff *et al.*, 2008). Nous n'avons donc pas pu exploiter la synténie avec les autres haplotypes de la région de *Bru1* pour combler l'espace de gauche. Nous avons dans un premier temps utilisé la technique de long-range PCR pour essayer d'amplifier le fragment d'ADN correspondant à cet espace. Dans un second temps, nous avons recherché s'il existait chez le sorgho une région orthologue à l'insertion présente sur l'haplotype cible. Enfin, nous avons utilisé des marqueurs flanquant l'espace de gauche pour tenter de combler cet espace.

### **B.1. Long-range PCR**

Nous avons essayé de combler cet espace de gauche en utilisant la technique de long-range PCR. Afin de mettre au point le protocole, nous avons défini des amorces tests sur le BAC CIR9O20 permettant d'amplifier 5 kb, 10 kb et 20 kb de séquence. Ces tests nous ont permis d'amplifier un fragment PCR de 10 kb aussi bien à partir d'ADN génomique qu'à partir d'ADN de BAC. Des amorces sens sur l'extrémité du BAC CIR9O20 et des amorces antisens sur l'extrémité du BAC 22M06 ont été définies le plus près possible de l'espace de gauche. Nous n'avons obtenu aucune amplification à partir de l'ADN génomique du cultivar de référence R570. Ces résultats suggèrent donc que cet espace de la carte physique de



**Figure.2.4. Visualisation du résultat du BLASTN avec le programme Circos obtenu d'après les séquences nucléotidiques des gènes présents dans l'insertion de l'haplotype cible et la séquence du génome total de sorgho disponible dans les bases de données.**

Les chromosomes du sorgho sont notés Sb1 à Sb10.

Seuls les alignements de plus de 100 nucléotides sont représentés avec une e-value minimum de 1E-05.



l'haplotype cible est supérieur à 10 kb.

## **B.2. Recherche d'une région orthologue chez le sorgho correspondant à l'insertion sur l'haplotype cible**

Pour réaliser la cartographie génétique fine du cultivar R570, une des stratégies a été d'exploiter les relations synténiques entre le sorgho et la canne à sucre (Le Cunff *et al.*, 2008), le sorgho étant l'espèce présentant les meilleures relations synténiques avec la canne à sucre.

Dans la même optique, nous avons essayé d'identifier la région du sorgho orthologue à l'insertion présente sur l'haplotype cible. Pour réaliser ce travail, nous avons créé un fichier avec toutes les séquences des gènes présents dans l'insertion sur le contig 22M06\_CIR12E03 : deux gènes codant pour des S/T kinases, un gène codant une CHP (Conserved Hypothetical Protein), un fragment de gène codant pour une NADP-dependent malic enzyme et un gène codant pour une zinc knuckle family protéine. Ces séquences nucléotidiques ont été alignées avec la séquence totale du génome du sorgho.

Les résultats sont présentés dans la figure 2.4. Seuls les alignements de plus de 100 nucléotides avec une e-value minimum de 1 E-05 sont représentés sur la figure. Les deux S/T kinases et la CHP n'ont pas présenté de forte homologie avec des séquences du génome du sorgho. Les meilleurs alignements pour les deux S/T kinases ne dépassent pas 30 nucléotides. La séquence codant pour la NADP-dependent malic enzyme présente quant à elle des homologies avec des séquences sur le chromosome 3 (95.2 % d'identité nucléotidique) et le chromosome 9 du sorgho (95.8 % d'identité nucléotidique). Les régions des chromosomes 9 et 3 du sorgho correspondent à un gène codant pour une NADP-dependent malic enzyme. Le gène codant pour une zinc knuckle family protéine présente des homologies avec plusieurs régions du sorgho situées sur plusieurs chromosomes. Le meilleur alignement se situe sur le chromosome 8 du sorgho (75.2 % d'identité nucléique) dans une région non annotée. Ce gène présente aussi une homologie avec une séquence sur le chromosome 3 du sorgho (73.2 % d'identité nucléique), mais pas dans la même région que la séquence codant pour la NADP-dependent malic enzyme. Les résultats obtenus ne nous permettent donc pas d'identifier clairement une région du sorgho orthologue à l'insertion.



### B.3. Criblage des banques BAC HindIII

#### B.3.1 Côté droit de l'espace de gauche

Nous avons défini des sondes à l'extrémité du BAC 22M06 à partir des données de séquençage. Très peu de sondes utilisables avaient été obtenues dans cette région. En effet, l'annotation de l'extrémité 5' du BAC 22M6 a révélé la présence de nombreux éléments transposables (les quatorze premiers kb du BAC) (Figure 2.3). Pour résoudre ce problème, nous avons défini deux sondes (TG22M6 - 643 pb - et TG22M6\_2 - 649 pb) dans une région intergénique d'environ 4 kb entre la fin de la région comprenant les éléments transposables et le gène 15 codant pour la S/T kinase. Il a été très difficile de définir des sondes car la région est très riche en GC. Les résultats n'ont pas pu être exploités car les sondes utilisées s'hybrident sur un nombre trop important de BAC (sondes polymorphes), ce qui rendait la lecture des filtres impossible.

#### B.3.2 Côté gauche de l'espace de gauche

Le BAC CIR9O20 présente à son extrémité un gène codant pour un fragment de NADP-dépendant D-sorbitol-6-phosphate déshydrogénase (*S6PDH*, gène 11a) interrompu par le site de clonage du BAC. Ce gène est aussi présent chez les autres haplotypes hom(é)ologues. Nous supposons que l'exon 1 du gène se trouve dans l'espace de gauche. En alignant les séquences des gènes *S6PDH11a* des différents haplotypes hom(é)ologues, nous avons dérivé une sonde dans un domaine conservé de l'intron 1 (sonde Sc1001 - 625 pb). Les banques BAC ont été criblées avec cette sonde : 8 clones BAC ont été identifiés sur la banque Clemson et 3 sur la banque CIRAD. Sur ces 11 BAC identifiés, 8 clones BAC étaient déjà connus comme n'appartenant pas à l'haplotype cible. Trois nouveaux clones BAC ont donc été identifiés : 234F20, CIR38O19 et CIR42E15.

L'haplotype cible présente deux copies du gène codant pour la *S6PDH11a* : une séquence tronquée sur le BAC CIR9O20 (CDS 11.a-1) et une séquence complète sur le contig



22M06\_CIR12E03 (CDS 11.a-2), en bordure de l'insertion présente sur l'haplotype cible (Figure 2.3). Pour différencier les BAC correspondant à ces différentes régions, nous avons, lors de notre vérification par amplification PCR, utilisé des marqueurs couvrant l'ensemble de la région : 9O20D11/RsaI\*, ER22M06\*, R12H16\* et F12H16. Les trois BAC ont été amplifiés seulement par le marqueur PCR ayant servi à faire la sonde Sc1001. Les fingerprint/profils de digestion de ces trois BAC n'ont présenté aucune bande commune avec les BAC connus de l'haplotype cible mais il est possible que ces BAC ne chevauchent ni le BAC CIR9O20 ni le BAC 22M06 ou que très légèrement (un chevauchement de quelques nucléotides). Les BES des trois BAC ont été séquencées. Les alignements des BES avec les BAC hom(é)ologues de la région de *Bru1* ont montré que :

- Le BAC 234F20 issu de la banque Clemson correspond vraisemblablement à un autre haplotype hom(é)ologue (haplotype VI ; Figure 2.3) car une des BES présente 100 % d'identité de séquence sur un alignement de 525 pb avec le BAC 135P16.

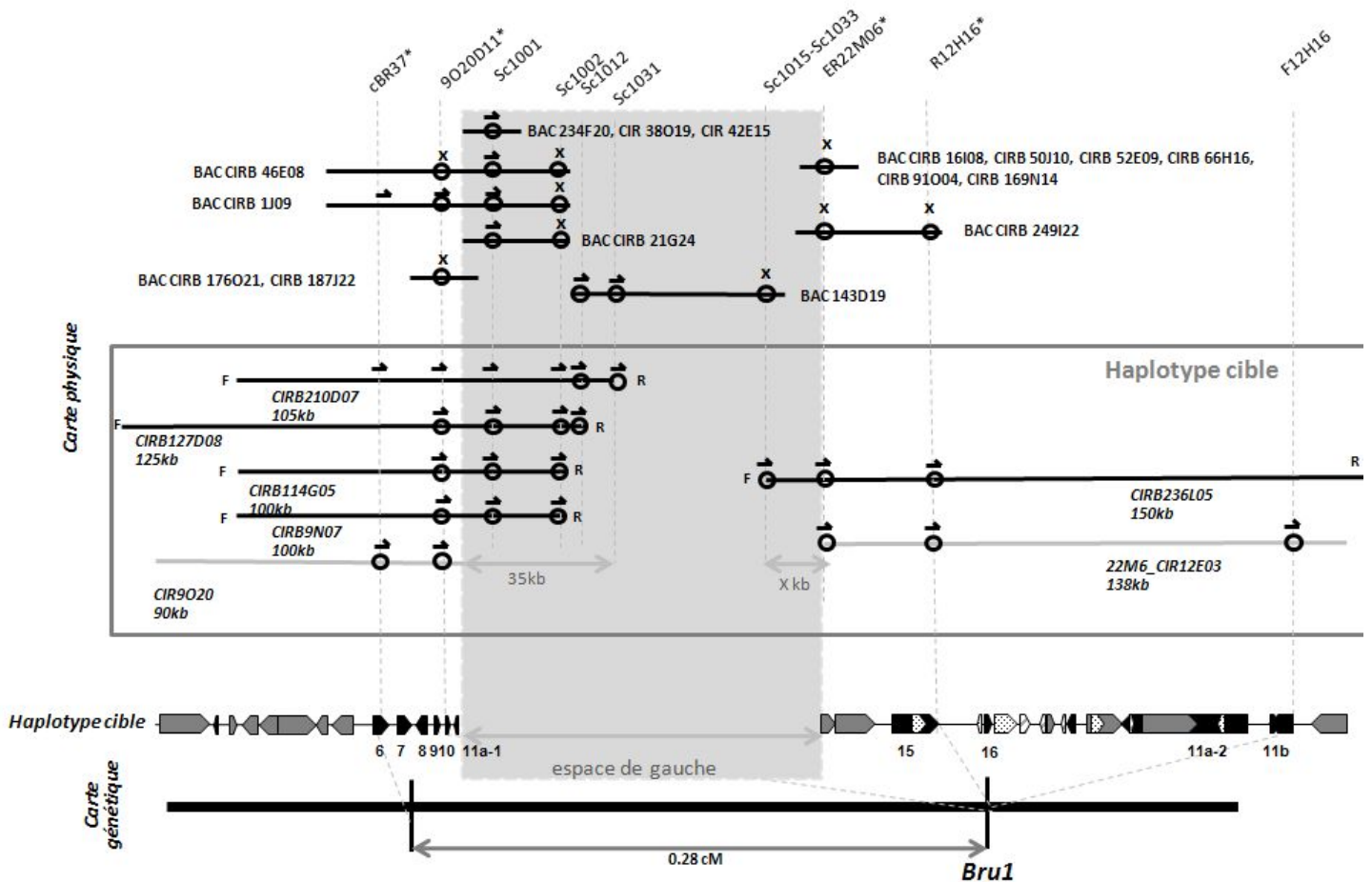
- L'extrémité sens (extrémité F) du BAC CIR38O19 présente 94% d'identité de séquence avec le BAC 53A11, au niveau d'un domaine intergénique situé en aval du gène 5 (Figure 2.3). L'extrémité antisens (extrémité R) ne présente aucune homologie avec les séquences connues de la région de *Bru1*.

- L'extrémité F du BAC CIR42E15 présente 98 % d'identité avec un LTR-rétrotransposon du BAC 53A11. Les deux plus forts pourcentages d'identité pour l'extrémité R ont été observés avec deux LTR-rétrotransposons de la famille Rhum situés à l'extrémité du contig 22M06\_CIR12E03 et du BAC 253G12 (83.8 % et 83.6 % d'identité respectivement) (Figure 2.3).

D'après les profils fingerprint et l'analyse des séquences des BES, aucun de ces 3 clones BAC n'appartient à l'haplotype cible.

### **B.4. Criblage de la banque BamHI**

La banque BAC CIRAD BamHI a été analysée avec 4 sondes définies précédemment : 9O20D11, Sc1001, ER22M06 et R12H16 (Figure 2.5). Les sondes 9O20D11, R12H16, et



**Figure.2.5. Carte physique et génétique de l'espace de gauche avec l'annotation des BACs déjà identifiés avant la thèse.**

Les traits horizontaux gris représentent les BACs déjà identifiés avant ma thèse. Les traits horizontaux noir représentent les nouveaux BACs identifiés. Le cercle noir représente que la sonde s'est hybridée sur le BAC. La demi-flèche noire correspond à une amplification PCR du marqueur sur le BAC. La croix noire n'indique aucune amplification PCR sur le BAC. Les marqueurs haplotypes-spécifiques sont indiqués par un astérique (\*).

Les gènes sont indiqués par des boîtes noires, LTR-rétrotransposons par des boîtes grises (le temps d'insertion est indiqué à l'intérieur), non-LTR rétrotransposons par des boîtes grises striées, transposons par des boîtes grises en pointillées et les « protéinase inhibitor » par des losanges blancs.

ER22M06 ont été choisies car elles hybridaient très peu de clones BAC sur la banque Clemson et les marqueurs PCR correspondants sont spécifiques de l'haplotype cible. La sonde Sc1001 avait été définie pour identifier des BAC dans l'espace de gauche.

La construction de la nouvelle banque BAC BamHI s'est réalisée en deux étapes. Une première banque de 40 000 clones a été construite puis elle a été étendue à 119 040 clones. Le criblage a donc été réalisé en deux étapes.

#### **B.4.1. Criblage de la première banque BamHI constituée de 40 000 clones**

Le premier criblage de la banque nous a permis d'identifier des nouveaux clones BAC (Figure 2.5)

- 3 BAC allumés avec la sonde 9O20D11 : CIRB1J09, CIRB9N07 et CIRB46E08
- 4 BAC allumés avec la sonde Sc1001, dont 3 BAC en commun avec la sonde 9O20D11 : CIRB1J09, CIRB9N07, CIRB46E08, et CIRB21G24
- 5 BAC allumés avec la sonde ER22M06 : CIRB16I08, CIRB50J10, CIRB52E09, CIRB66H16 et CIRB91O04
- aucun BAC allumé avec la sonde R12H16

Ces 12 BAC ont été amplifiés par PCR avec les amorces correspondant aux marqueurs 9O20D11/RsaI\*, R12H16\*, Sc1001 et ER22M06\*. Nous n'avons obtenu aucune amplification avec les BAC allumés par la sonde ER22M06. Les BAC CIRB21G24 et CIRB46E08 n'ont été amplifiés par aucun des marqueurs PCR et leur profil fingerprint ne présentait aucune bande commune avec les BAC CIR9O20 et 22M06. Ces BAC sont donc exclus de l'analyse.

Les BAC CIRB1J09 et CIRB9N07 ont été amplifiés par les marqueurs 9O20D11/RsaI\* et Sc1001. Le profil fingerprint du CIRB9N07 a révélé six bandes communes avec le BAC CIR9O20. Le fingerprint du CIRB1J09 a présenté quant à lui une seule bande commune avec le CIR9O20

Les BES des BAC CIRB1J09 et CIRB9N07 ont été séquencées. Les séquences ont été alignées avec les séquences des BAC de la région de *Brul* :





- CIRB1J09 : l'extrémité F ne s'aligne avec aucune séquence des haplotypes hom(é)ologues de la région de *Bru1*. Son extrémité présente une homologie de séquence avec un élément transposable de sorgho. Son extrémité R ne présente aucune homologie avec aucune séquence connue. Les données PCR nous ont montré que ce BAC était amplifié par un marqueur haplotype spécifique (9O20D11/RsaI\*) mais le résultat des BES ne nous permet pas d'ancrer ce BAC sur le BAC CIR9O20 et le profil de digestion du CIRB1J09 ne nous montre pas qu'il recouvre totalement le BAC CIR9O20 (une seule bande commune). Afin de comprendre cette situation, nous avons utilisé un autre marqueur spécifique, cBR37\*, présent sur le CIR9O20 et distant de 25 kb par rapport à l'espace de gauche. La bande amplifiée présente la taille attendue. Ce résultat implique que le BAC CIRB1J09 recouvre le CIR9O20 jusqu'au marqueur cBR37\*, soit un minimum de 25 kb. Or, le profil de digestion *in silico* du BAC CIR9O20 par BamHI nous a indiqué que, dans ces 25 kb, il existe deux sites de restriction qui génèrent deux fragments : un de 6,5 kb et un autre de plus de 21 kb (CIR9O20 est cloné avec l'enzyme HindIII, donc le BAC s'arrête sur un site HindIII). La bande commune révélée par les fingerprint ne correspond pas à cette bande de 6,5 kb. Les résultats obtenus ne sont donc pas cohérents et nous n'arrivons pas à les expliquer. Ce BAC a été mis de côté pour la suite de l'analyse.

- CIRB9N07 : l'extrémité F s'aligne à 100% sur le CIR9O20 à environ 22 kb de l'extrémité 5' de ce BAC, dans un domaine intergénique. L'extrémité R s'aligne sur quatre haplotypes hom(e)ologues dans un domaine intergénique juste en aval du gène *S6PDH11a-1*. Les profils de digestion nous indiquent que ce BAC couvrirait l'espace de gauche d'environ 30 kb par rapport au CIR9O20. Un marqueur PCR a été dérivé de son extrémité F et a permis de confirmer l'ancrage de ce BAC sur le BAC CIR 9O20. La sonde Sc1002 a été dérivée de l'extrémité R pour cribler la banque BAC BamHI de 40 000 clones BAC. Dix-sept clones BAC ont été identifiés, dont quatre BAC connus (CIRB1J09, CIRB9N07, CIRB21G24 et CIRB46E08). Seul le BAC CIRB9N07 a été amplifié par PCR avec l'amorce Sc1002.

Le criblage des 40 000 premiers clones de la banque BamHI a permis d'identifier un seul nouveau BAC appartenant à l'haplotype cible : CIRB9N07. Ce BAC, d'environ 100 kb, couvre partiellement l'espace de gauche (estimé à environ 30 kb).



#### **B.4.2. Criblage de la banque BamHI étendue à 119 040 clones**

Afin d'augmenter le taux de recouvrement de la banque BAC BamHI (40 000 clones pour un recouvrement de 0,8 X), nous avons continué de produire des clones BAC pour atteindre un nombre de 119 040 clones pour un recouvrement de 1,6 X génome du cultivar R570.

Le criblage des 80 000 nouveaux clones s'est effectué avec les mêmes sondes que celles utilisées précédemment : 9O20D11, ER22M06, R12H16 et Sc1002 définie à l'extrémité du BAC CIRB9N07. Quatre-vingt deux nouveaux BAC ont été identifiés (Figure 2.5) :

- la sonde 9O20D11 s'hybride sur 4 clones BAC : CIRB114G05, CIRB127D08, CIRB176O21 et CIRB187J22.
- la sonde Sc1002 s'hybride sur 73 clones BAC, dont 2 BAC en commun avec la sonde 9O20D11 (CIRB114G05 et CIRB127D08).
- la sonde ER22M06 s'hybride sur 3 clones BAC : CIRB169N14, CIRB236L05 et CIRB249I22.
- la sonde R12H16 s'hybride sur 2 clones BAC : CIRB236L05 et CIRB249I22. Ces BAC ont également été révélés par la sonde ER22M06.

Tous ces BAC ont été analysés par amplification PCR avec les amorces PCR dérivées des sondes. Sur les 82 BAC identifiés, seuls 3 BAC ont été validés par PCR (CIRB114G05, CIRB127D08 et CIRB236L05).

Le BAC CIRB114G05 a été identifié avec les sondes 9O20D11 et Sc1002. Il a été validé par PCR avec le marqueur diagnostique 9O20D11/RsaI\* et le marqueur Sc1002. Les séquences BES présentent 100 % d'identité avec les séquences BES du CIRB9N07. Ces données corroborent les résultats fingerprint qui montrent que le BAC possède exactement le même profil de digestion que le BAC CIRB9N07. La taille du CIRB114G05 est estimée à 100 kb. Ce BAC appartient à l'haplotype cible mais est identique au BAC CIRB9N07.

Le BAC CIRB127D08 a été identifié avec les sondes 9O20D11 et Sc1002. Il a été validé en PCR par les marqueurs 9O20D11/RsaI\* et Sc1002. D'après les données de fingerprint, il recouvre totalement le CIR9O20 et le CIRB9N07. Les séquences BES ne s'alignent sur aucun des BAC de l'haplotype cible. L'extrémité F ne présente aucune homologie avec des séquences connues disponibles dans les bases de données. L'extrémité R



de ce BAC présente en moyenne 98 % d'identité avec les quatre BAC hom(é)ologues, dans un domaine intergénique en aval du gène S6PDH11a-1. L'extrémité R du CIRB9N07 présente le même type d'alignement dans cette région, à environ 1 kb de l'extrémité R du CIRB127D08. Ce BAC, d'environ 125 kb, appartient à l'haplotype cible et avancerait d'environ 1 kb par rapport au CIRB9N07 et d'environ 31 kb par rapport au CIR9O20 sur la base des alignements des extrémités.

Le BAC CIRB236L05 a été identifié avec les sondes ER22M06, R12H16 et Sc1002. Il a été vérifié en PCR par les marqueurs ER22M06\* et R12H16\*. Nous avons aussi effectué des vérifications avec un marqueur PCR de l'espace de droite, F12H16. Le profil de digestion indique qu'il recouvre totalement le contig 22M06\_CIR12E03. Les BES de ce BAC ne présentent aucune homologie de séquence avec les séquences des BAC de l'haplotype cible. L'extrémité R présente des homologies de séquence avec des éléments transposables de canne à sucre appartenant à la famille des LTR-Rétrotransposons RLG-Colada (80 % d'identité nucléique). Sa taille est estimée à 150 kb. Le BAC CIRB236L05 appartiendrait donc à l'haplotype cible et permettrait d'avancer dans les deux espaces. Afin d'orienter le BAC, des marqueurs PCR ont été dérivés de ses extrémités. Le marqueur Sc1014 (extrémité R) amplifie le BAC CIRB286F09 qui couvre partiellement l'espace de droite (voir paragraphe C.3). L'extrémité R du BAC se situe dans le trou de droite et permet d'avancer dans cet espace. L'extrémité F, quant à elle, se situe dans l'espace de gauche. Son extrémité sera utilisée comme sonde (Sc1015) pour cribler les banques BAC.

Le criblage des 80 000 nouveaux clones de la banque BamHI a permis d'identifier trois nouveaux BAC appartenant à l'haplotype cible : deux BAC du côté gauche de l'espace de gauche (CIRB114G05 et CIRB127D08) et un BAC du côté droit de l'espace de gauche (CIRB236L05). Le CIRB114G05 est identique au CIRB9N07. Le CIRB127D08 couvre partiellement l'espace de gauche et avancerait d'environ 1 kb par rapport au CIRB9N07. Le BAC CIRB236L05 couvre partiellement les deux espaces de la carte physique de l'haplotype cible sur une distance indéterminée (Figure 2.5).

Une sonde (Sc1012 - extrémité R) a été dérivée de l'extrémité R du BAC CIRB127D08. Deux sondes (Sc1014 - extrémité R et Sc1015 - extrémité F) ont été dérivées des deux extrémités du CIRB236L05



### B.4.3. Marche sur chromosome à partir des trois banques BAC

La sonde Sc1012 (extrémité R du CIRB127D08) et la sonde Sc1015 (extrémité F du CIRB236L05) ont été utilisées pour cribler les trois banques BAC (Figure 2.5). Cent quatre vingt seize et sept clones BAC ont été respectivement identifiés avec les sondes Sc1012 et Sc1015. Le BAC 143D19 a été identifié par les deux sondes.

Les vérifications PCR (marqueurs 9O20D11/RsaI\*, Sc1002, Sc1012 et Sc1015) sur tous les BAC identifiés nous ont permis de retenir trois nouveaux BAC :

Le BAC CIRB210D07 a été révélé par la sonde Sc1012. Il a été vérifié en PCR par les marqueurs 9O20D11/RsaI\*, Sc1002 et Sc1012. Les profils de digestion confirment qu'il appartient à l'haplotype cible : toutes les bandes du CIRB9N07 sont retrouvées. Ces résultats sont confirmés avec les comparaisons de séquence des BES. L'extrémité F s'ancre à 100 % sur le CIR9O20, à la même position que l'extrémité F du CIRB9N07. L'extrémité R présente en moyenne 96 % d'identité avec les quatre BAC hom(é)ologues, au niveau d'un domaine intergénique en aval du gène S6PDH11a-1, dans la même région que l'extrémité R du CIR9N07 et que l'extrémité R du CIRB127D08. L'extrémité du CIRB210D07 se situerait donc à environ 3 kb de l'extrémité R du CIRB127D08 sur la base des homologies de séquence. Sa taille est estimée à environ 105 kb. Ce BAC appartient bien à l'haplotype cible et avancerait dans l'espace de gauche d'environ 4 kb par rapport au CIRB9N07 et d'environ 35 kb par rapport au CIR9O20 (Figure 2.5).

Le BAC 143D19 a été révélé par les sondes Sc1012 et Sc1015. Il a été vérifié par le marqueur Sc1012, mais n'est pas amplifié par la Sc1015. Le profil de digestion indique qu'il ne possède aucun fragment en commun avec les BAC CIRB210D07 et CIRB236L05. Les deux clones BAC bordant l'espace de gauche sont clonés avec l'enzyme BamHI tandis que le BAC 143D19 est cloné avec l'enzyme HindIII. La lecture des profils de digestion peut donc s'avérer difficile, surtout si le BAC143D19 chevauche de très peu ces deux BAC. La taille du clone BAC est estimée à 120 kb. La séquence BES de l'extrémité F ne s'aligne avec aucune séquence de la région de *Bru1* mais présente des homologies de séquence avec des éléments transposables de canne à sucre. La séquence BES de l'extrémité R présente 86 % d'identité avec un élément transposable du sorgho annoté comme un LTR-rétrotransposon. Pour l'instant, nous ne pouvons écarter l'appartenance de ce BAC à l'haplotype cible.





Le criblage des trois banques BAC disponibles avec ces trois sondes nous a donc permis d'identifier un BAC (CIRB210D07) appartenant à l'haplotype cible, qui couvre partiellement l'espace de gauche et avancerait d'environ 4 kb par rapport à l'extrémité du BAC CIRB9N07 (Figure 2.5). La sonde Sc1031 a été dérivée de l'extrémité R du BAC CIRB210D07. Le BAC 143D19 reste un candidat potentiel car notre méthode de vérification ne nous a pas permis de l'exclure.

La sonde Sc1015 ne s'est hybridée que sur 7 clones BAC sur les 330 000 clones de la banque et aucun des BAC révélé par cette sonde n'est amplifié par PCR, à l'exception du BAC CIRB236L05 (BAC où la sonde a été définie). Pour essayer de comprendre cette situation, nous avons défini une nouvelle amorce (Sc1033) à l'intérieur du marqueur Sc1015. Les résultats PCR nous indiquent que six des sept BAC sont amplifiés par l'amorce Sc1033 mais pas à la taille attendue (amplificat de 600 pb au lieu de 300 pb attendu). De plus, nous avons effectué un fingerprint de tous ces BAC puis celui-ci a été hybridé par la sonde Sc1033. Aucun des BAC ne possède la bande du BAC CIRB236L05 révélé par cette sonde. En parallèle, nous avons fait séquencer les BES des six BAC révélés par cette sonde (dont le BAC 143D19). Les séquences des BES ne nous permettent d'ancrer aucun de ses six BAC dans la région de *Bru1*.

Afin de continuer notre marche sur chromosome, nous avons criblé les 3 banques BAC avec la sonde Sc1031 (extrémité du BAC CIRB210D07). Quarante BAC ont ainsi été identifiés. Huit de ces 40 BAC étaient déjà connus pour appartenir aux autres haplotypes hom(é)ologues. Les 32 BAC restant ont été analysés par PCR avec les marqueurs Sc1002, Sc1012 et Sc1031. Seul le BAC 143D19 a été vérifié par le marqueur Sc1031 (ce BAC avait déjà été identifié puis amplifié par le marqueur Sc1012). Ce criblage ne nous a donc pas permis d'identifier de nouveaux clones BAC appartenant à l'haplotype cible.

### **B.5. Conclusion sur la carte physique de l'espace de gauche**

Le criblage des banques BAC nous a permis d'identifier cinq nouveaux BAC appartenant à l'haplotype cible (Figure 2.5): CIRB9N07, CIRB114G05, CIRB127D08, CIRB210D07 et le CIRB236L05. Quatre BAC (CIRB9N07, CIRB114G05, CIRB127D08,



CIRB210D07) ont été identifiés comme couvrant partiellement (environ 35 kb) le côté gauche de l'espace de gauche et un BAC (CIRB236L05) comme couvrant partiellement le côté droit de l'espace de gauche. Le BAC CIRB236L05 couvre également partiellement l'espace de droite (taille indéterminée). La stratégie employée de validation d'un BAC pour l'appartenance à l'haplotype cible ne nous a pas permis de conclure au sujet du BAC 143D19, qui reste donc un candidat potentiel.

Les derniers résultats de criblage ne nous ayant pas permis d'identifier de nouveaux clones BAC pouvant appartenir à l'haplotype cible, nous avons décidé de séquencer les BAC CIRB210D07, CIRB236L05 et 143D19 (Centre National de Séquençage - Génoscope, Evry, France). D'après nos analyses, les BAC CIRB210D07 et CIRB236L05 sont les BAC qui avanceraient le plus loin de chaque côté de l'espace de gauche. Seul le séquençage du BAC 143D19 nous permettra soit de valider ce BAC comme appartenant à l'haplotype cible soit de l'écarter de l'analyse.

### **C. Cartographie physique pour compléter l'espace de droite**

#### **C.1. Eliminer l'espace de droite par réduction de la région cible**

Nous avons tenté d'éliminer l'espace de droite de la région cible en tentant d'identifier de nouvelles recombinaisons plus proches du gène *Bru1* au niveau de l'extrémité F du BAC CIR12E03 (Figure 2.3). Nous avons défini six nouveaux marqueurs à l'extrémité de ce BAC, dont cinq nouvelles sondes situées dans un domaine intergénique entre deux éléments transposables. Ces sondes ont été analysées sur un blot de tri Résistants/Sensibles digéré par cinq enzymes différentes. Les résultats obtenus nous montrent que ces cinq sondes sont très polymorphes et donc inexploitable pour notre analyse. Le sixième marqueur est un marqueur PCR haplotype-spécifique (Sc1003\*) que nous avons réussi à définir à l'extrémité du BAC CIR12E03, au niveau 3' du gène codant pour la *S6PDH11b*, à proximité du marqueur F12H16. Ce marqueur a pu être cartographié, il coségrège avec le gène *Bru1* mais il ne nous a pas permis de réduire la région cible



## C.2. Criblage des banques HindIII

A partir de la carte physique décrite au paragraphe 1-E, nous avons entrepris de combler l'espace de droite de l'haplotype cible en utilisant la synténie existante entre les différents haplotypes hom(é)ologues pour définir de nouvelles sondes et ainsi cribler les banques BAC. Les données d'annotation et de comparaison des BAC hom(é)ologues de la région de *Bru1* (Figure 2.3) ont montré qu'un gène correspondant à une glucanase était présent sur les cinq haplotypes hom(é)ologues au niveau de l'espace de droite de l'haplotype cible. Au vue de la forte colinéarité des gènes entre les BAC hom(é)ologues, il était fort probable que ce gène soit présent dans l'espace de droite. Nous avons donc défini, à partir de l'alignement nucléotidique de toutes les glucanases présentes sur les différents haplotypes de la région de *Bru1*, une sonde ciblant l'exon 3 du gène (Glu2 - 688 pb), une sonde ciblant l'intron 1, l'exon 2 et l'intron 2 du gène (Glu3 - 286 pb). Ces sondes ont permis d'identifier des nouveaux BAC :

- 58 BAC de la banque Clemson et 8 BAC de la banque CIRAD HindIII ont été révélés par la sonde Glu2
- 22 BAC de la banque Clemson et 27 BAC de la banque CIRAD HindIII ont été révélés par la sonde Glu3.

La difficulté majeure rencontrée pour cette espace était l'absence de marqueurs spécifiques de l'haplotype cible. Cette contrainte intervient principalement dans les étapes de vérifications PCR des BAC nouvellement identifiés. A cette étape du criblage des banques HindIII, nous ne disposions pas encore de marqueurs haplotypes spécifiques. Nous considérons donc qu'un BAC n'était validé par PCR que s'il avait été amplifié par au moins deux marqueurs PCR. Les BAC identifiés par les sondes Glu2 et Glu3 ont été analysés par PCR avec les marqueurs flanquants de l'espace de droite (F12H16, F164H22) et les marqueurs PCR séquences des BAC hom(é)ologues de la région de *Bru1* indiquent que :

- les extrémités F du BAC 63P17 et R du BAC 263N18 présentent 100 % d'identité avec le BAC 135P16. Ces deux BAC n'appartiennent donc pas à l'haplotype cible mais à l'haplotype VI de la région de *Bru1*.

- l'extrémité F du BAC 181A13 présente 100 % d'identité avec le BAC 15N23. Ce BAC n'appartient donc pas à l'haplotype cible mais à l'haplotype I de la région de *Bru1*.



- les BES du BAC 98E04 ne présentent aucune homologie avec des BAC de la région de *Bru1*. De plus, le fingerprint du BAC 98E04 ne présente aucune bande commune avec les BAC connus de l'haplotype cible. Ce BAC n'appartiendrait donc pas à l'haplotype cible et est écarté de l'analyse.

(2) six BAC ont été amplifiés avec les marqueurs PCR F12H16 et Glu3 : 204L10, 218A09, 253P15, CIR26M20, CIR30H01 et CIR43N23. Aucune homologie des BES avec les séquences des BAC de la région de *Bru1* n'a été identifiée. Aucun des fingerprint de ces six BAC ne présente de similarité avec les fingerprint des BAC connus de l'haplotype cible.

Le criblage des deux banques HindIII ne nous a pas permis d'identifier de nouveau BAC appartenant à l'haplotype cible et ainsi de progresser dans l'espace de droite.

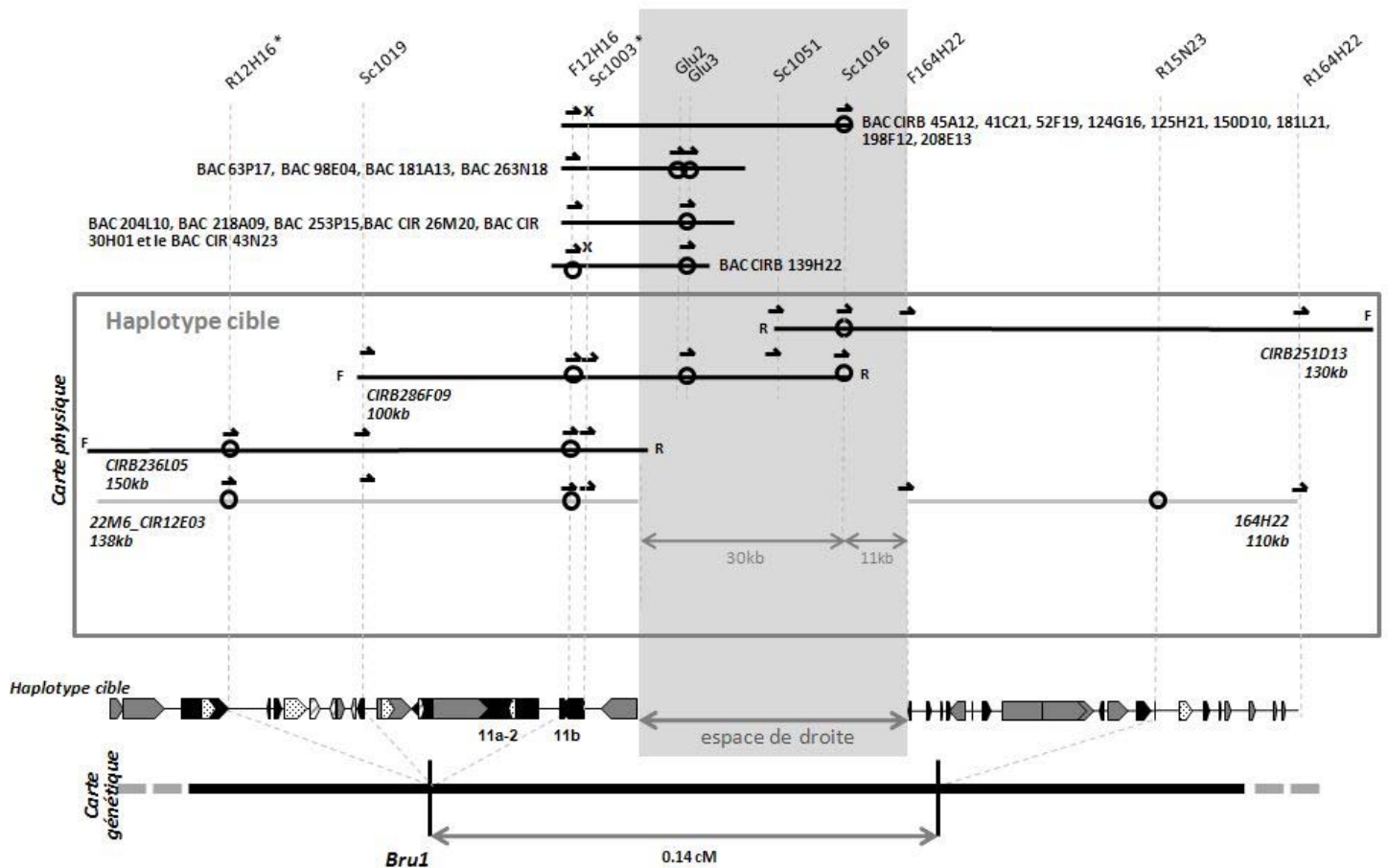
### **C.3. Criblage de la banque BamHI**

La banque BAC CIRAD BamHI a été criblée avec deux sondes, F12H16 et Glu3. La sonde F12H16 avait déjà permis de cribler les banques HindIII et s'était hybridé sur très peu de BAC. Ce criblage nous a permis d'identifier 18 et 115 nouveaux clones BAC, révélés par les sondes F12H16 et Glu3 respectivement. Deux clones BAC ont été révélés par les deux sondes (CIRB139H22 et CIRB286F09). A cette étape du criblage, nous avons pu développer un marqueur PCR haplotype spécifique, Sc1003\*, présenté dans le paragraphe C.1.

Les analyses par amplification PCR avec les marqueurs F12H16, Sc1003\*, Glu3 et F164H22 (Figure 2.6) ont montré que deux BAC étaient amplifiés par au moins deux marqueurs PCR. Ils correspondent aux deux BAC allumés par les deux sondes :

- le BAC CIRB139H22 a été amplifié par les marqueurs PCR F12H16 et Glu3 mais pas par le marqueur haplotype spécifique Sc1003\*. Ce BAC a donc été exclu pour la suite de l'analyse.

- le BAC CIRB286F09 est amplifié avec les marqueurs F12H16, Sc1003\* et Glu3 mais pas avec le marqueur PCR F164H22. Les profils de digestion confirment que ce clone appartient à l'haplotype cible : il possède des bandes communes avec le BAC 22M06 et toutes



**Figure.2.6. Carte physique et génétique de l'espace de droite avec l'annotation des BACs déjà identifiés avant la thèse.**

Les traits horizontaux gris représentent les BACs déjà identifiées avant ma thèse. Les traits horizontaux noir représentent les nouveaux BACs identifiés. Le cercle noir représente que la sonde s'est hybridée sur le BAC. La demi-flèche noire correspond à une amplification PCR du marqueur sur le BAC. La croix noire n'indique aucune amplification PCR sur le BAC. Les marqueurs haplotypes-spécifiques sont indiqués par un astérisque (\*).

Les gènes sont indiqués par des boîtes noires, LTR-rétrotransposons par des boîtes grises (le temps d'insertion est indiqué à l'intérieur), non-LTR rétrotransposons par des boîtes grises striées, transposons par des boîtes grises en pointillées et les « protéinase inhibitor » par des losanges blancs.



les bandes du BAC CIR12E03. L'extrémité F du BES présente 100% d'identité de séquence avec le BAC 22M06\_CIR12E03 et s'ancore sur le BAC 22M06\_CIR12E03 à environ 66 kb en amont de l'espace de droite. L'extrémité R présente quant à elle 99% d'identité avec une région intergénique située 20 kb en amont du gène codant pour la glucanase d'un des BAC hom(é)ologues (BAC 253G12). La taille du BAC est estimée à 90 kb. Un marqueur PCR (Sc1019) a été défini à l'extrémité F du BAC CIRB286F09. Ce marqueur nous a permis de confirmer l'ancrage du CIRB286F09 sur le BAC 22M06\_CIR12E03.

Le BAC CIRB286F09 appartient donc à l'haplotype cible et couvre partiellement l'espace de droite d'environ 20 kb. Pour poursuivre le criblage des banques, une sonde (Sc1016) a été dérivée de son extrémité R. Dans l'optique d'obtenir un marqueur spécifique de l'haplotype cible dans cet espace, nous avons essayé de dériver un marqueur CAPS (Cleaved Amplified Polymorphic Sequence) à partir de la sonde Sc1016. Cinq enzymes différentes ont été testées sur le produit PCR du BAC, sur l'ADNg de R570 et sur l'ADN des bulks résistant et sensible. Aucun marqueur spécifique n'a pu être défini.

### **C.4. Marche sur chromosome à partir des trois banques BAC**

La sonde Sc1016 s'est hybridée sur 242 clones BAC des banques Clemson et CIRAD BamHI. Ces clones BAC ont été vérifiés par PCR avec des couples d'amorces encadrant l'espace de droite et avec le couple d'amorces dérivé de la sonde Sc1016. Seuls les BAC présentant au moins une amplification avec deux marqueurs PCR, dont la Sc1016, ont été retenus, soit dix BAC au total (Figure 2.6). Neuf BAC ont été amplifiés par PCR avec les marqueurs F12H16 et Sc1016 mais ont été exclus de l'analyse car ils n'étaient pas amplifiés par le marqueur spécifique de l'haplotype cible (Sc1003\*). Le BAC CIRB251D13 a été amplifié en PCR par les marqueurs Sc1016 et F164H22. Son profil de digestion nous indique qu'il possède des bandes communes avec le CIRB286F09 et toutes les bandes du BAC 164H22. Il recouvrirait donc entièrement ce BAC. Pour confirmer ce résultat, nous avons réalisé une PCR avec un marqueur (R164H22) ciblant l'autre extrémité du BAC 164H22. Nous avons obtenu une amplification de même taille sur le BAC CIRB251D13. Les BES du BAC CIRB251D13 ne présentent aucune homologie avec les autres BAC séquencés de l'haplotype cible. Par contre, l'amorce PCR dérivée de son extrémité R (marqueur Sc1051) a



amplifié à la taille attendue le BAC CIRB286F09 (Figure 2.6). La taille du BAC est estimée à 140 kb. Le BAC CIRB251D13 appartient à l'haplotype cible. Il recouvrirait partiellement le BAC CIRB286F09 et entièrement le BAC 164H22. Ce BAC, avec le BAC CIRB286F09, permettraient de combler l'espace de droite présent sur l'haplotype cible.

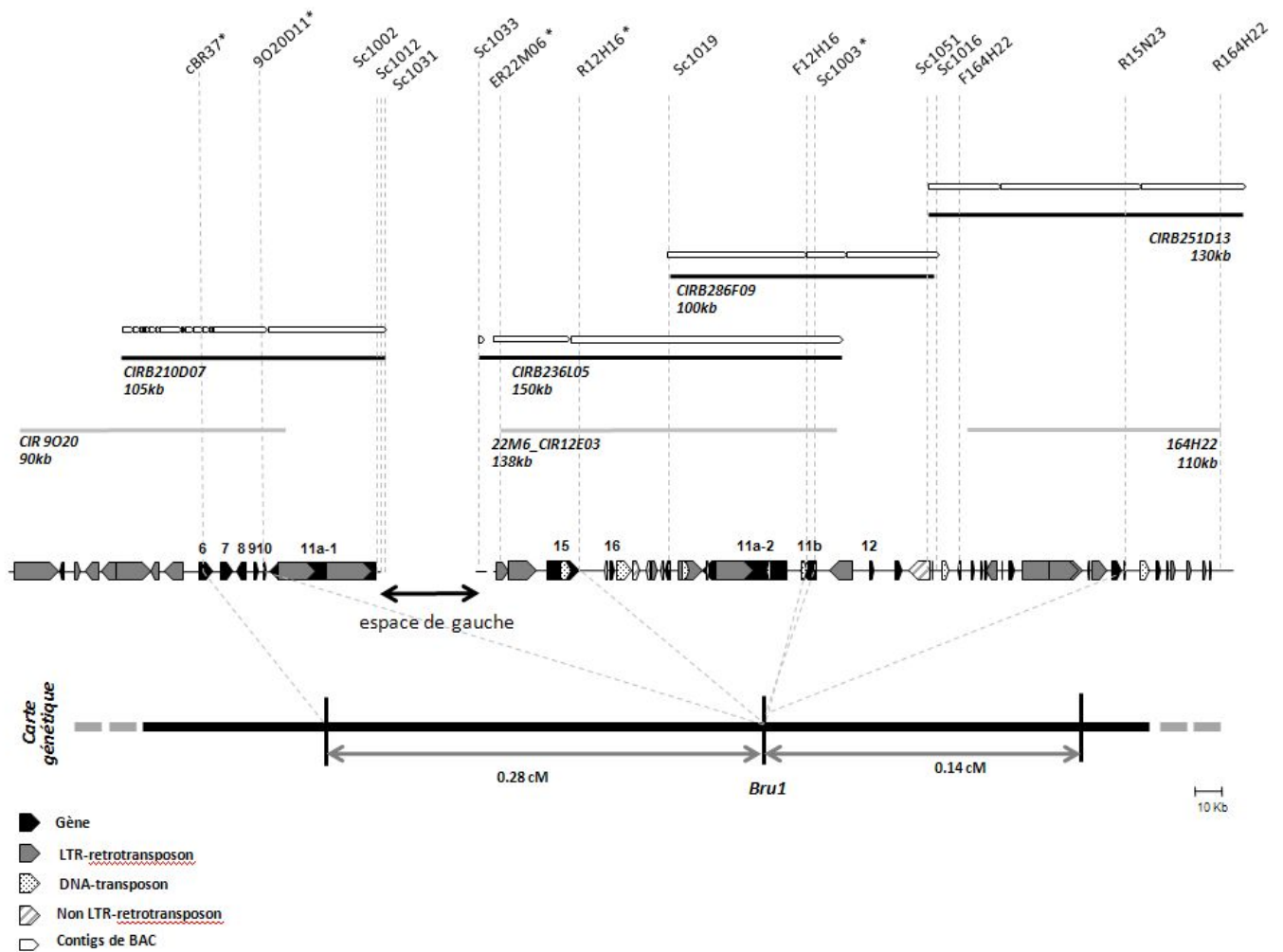
### **C.5. Conclusion sur la carte physique de l'espace de droite**

La stratégie basée sur la synténie existant entre les différents haplotypes hom(é)ologues de la région de *BruI* nous a donc permis de combler l'espace de droite de la carte physique de l'haplotype cible. En effet, deux nouveaux clones BAC ont été identifiés et validés comme appartenant à l'haplotype cible : CIRB286F09 et CIRB251D13. Ces deux BAC permettraient de combler l'espace de droite de l'haplotype cible. Ils ont été envoyés à séquencer (Centre National de Séquençage - Génoscope, Evry, France) afin de déterminer le contenu en gènes de cet espace.

### **D. Séquençage et annotation des nouveaux BAC de l'haplotype cible**

Deux clones BAC recouvrant partiellement l'espace de gauche (CIRB 210D07 et CIRB 236L05) et deux clones BAC recouvrant totalement l'espace de droite (CIRB 286F09 et CIRB 251D13) ont été séquencés, ainsi que le BAC 143D19 qui n'avait pas pu être écarté de l'analyse par notre méthode de validation. Les résultats de l'annotation sont présentés sur la figure 2.7.

La séquence du BAC 143D19 est composée de quatre contigs représentant 117 kb. La séquence de ces quatre contigs ne présente aucune homologie avec les séquences connues des BAC de la région de *BruI*. L'annotation du BAC 143D19 révèle la présence d'un seul élément transposable de type LTR rétrotransposon et de 14 gènes dont 10 gènes annotés comme codant pour des « Hypothetical Proteins » (HP). Aucun des autres gènes ne correspond à des gènes présents dans la région de *BruI*. Ce résultat est surprenant car ce BAC



**Figure 2.7. Annotation des nouveaux BACs identifiés appartenant à l'haplotype cible.**  
Les \* indiquent les marqueurs haplotypes-spécifiques.

a été allumé par des sondes ciblant le gène codant pour la *S6PDH*. Nous nous attendions donc à identifier au moins un gène *S6PDH* (11a ou 11b) sur la séquence. D'après ces résultats, ce BAC n'appartiendrait pas à l'haplotype cible et n'a donc pas été retenu pour la suite.

La séquence du BAC CIRB210D07 est composée de 19 contigs de 485 pb à plus de 42 kb, représentant 104 kb de séquence totale. Les séquences des BES du CIRB210D07 ont été utilisées pour orienter certains contigs. L'extrémité F du CIRB210D07 s'ancre à plus de 21 kb du début du BAC CIR9O20. Ce BAC chevauche le CIR9O20 de plus de 65 kb et rentre dans l'espace de droite de plus de 34 kb (extrémité R du CIRB210D07). L'assemblage des contigs a été vérifié par comparaison avec la séquence chevauchante du CIR9O20. Tous les contigs sont adjacents et s'ancrent sur le CIR9O20, avec 99,5 % d'identité nucléotidique en moyenne (Figure 2.7). Un fragment de 2 kb du CIRB210D07 ne présente aucune homologie de séquence avec le CIR9O20 alors que celui-ci a été assemblé sur la partie chevauchante du CIR9O20. Pour vérifier l'assemblage de ce fragment, des amorces encadrant ce fragment ont été testées sur les BAC CIR9O20, CIRB9N07, CIRB127D08, CIRB210D07 et sur l'ADNg de R570. Aucune amplification PCR n'a été obtenue pour le BAC CIR9O20 mais une amplification de la taille attendue a été observée pour les BAC de la banque BamHI et R570. L'assemblage du BAC CIRB210D07 semble donc être correct. Afin de vérifier la qualité de la séquence, nous avons recherché le nombre de SNPs (Single Nucleotide Polymorphism) au niveau des 65 kb de chevauchement entre les deux séquences de BAC, grâce à l'outil SNIPlay ([www.sniplay.cirad.fr](http://www.sniplay.cirad.fr)) qui permet de rechercher les SNPs à partir d'un alignement de séquences nucléotidiques. Dix-huit indels (insertion / délétion) et 30 SNPs ont été identifiés entre les BAC CIRB210D07 et CIR9O20. Les 18 indels sont situés dans des zones de répétition du même nucléotide (poly A, poly T, poly C ou poly G) et peuvent donc s'expliquer par des erreurs de séquençage. Nous avons 30 SNP sur 65 kb de chevauchement soit 1 SNP tous les 2 kb environ, ce qui reste une valeur élevée pour des BAC homologues. L'annotation des 34 369 pb du BAC CIRB210D07 (Figure 2.7) nous a permis de retrouver le début du gène 11a (exon 1 et une partie de l'intron 1) qui était tronqué sur le BAC CIR9O20. Ce gène est complet sur le contig CIR9O20\_CIRB210D07, avec une taille totale de 34 845 pb. L'intron 1 de ce gène représente 32 kb de séquence, avec deux LTR rétrotransposons insérés, dont un appartenant à la superfamille des Copia RLC-Rhum (15 114 pb) et un autre



appartenant à la superfamille des Gypsy (15 107 pb).

La séquence du BAC CIRB236L05 est composée de trois contigs représentant au total 142 kb. Deux des contigs permettent de recouvrir totalement le contig 22M06\_CIR 12E03. Le premier contig recouvre le BAC 22M06\_CIR12E03 de 30 215 pb et permet d'avancer dans l'espace de gauche de 1 821 pb (Figure 2.7). Le deuxième contig recouvre le 22M06\_CIR12E03 de 107 457 pb et permet d'avancer de 520 pb dans l'espace de droite (extrémité R du CIRB236L05). Le dernier contig de 2 000 pb contient l'extrémité F du CIRB236L05 et donc se situe dans l'espace de gauche. Vingt-deux indels et aucun SNP différencient les séquences chevauchantes de ces deux BAC. Les 22 indels sont situés dans des zones de répétition du même nucléotide (poly A, poly T, poly C ou poly G), comme pour le séquençage du BAC CIRB210D07. L'annotation des séquences du BAC CIRB236L05 (Figure 2.7) permettant d'avancer dans les deux espaces n'a révélé la présence d'aucun gène. Le peu de séquences exploitables - 4 kb pour l'espace de gauche et 520 pb pour l'espace de droite - ne nous ont pas permis d'identifier de nouveaux gènes.

La séquence du BAC CIRB286F09 est composée de trois contigs représentant au total 101 kb de séquence. Les séquences BES du CIRB286F09 ont été utilisées pour orienter les contigs. L'extrémité F du CIRB286F09 s'ancre à 66 498 pb du début du contig 22M06\_CIR12E03, au niveau du gène codant pour la zinc knuckle family protéine. La comparaison des séquences des contigs obtenus avec les séquences des BAC connus de l'haplotype cible indique que le BAC CIRB286F09 chevauche le contig 22M06-CIR12E03 de 71 193 pb et rentre dans l'espace de droite de 29 727 pb (extrémité R du CIRB 286F09). Vingt-et-un indels et 4 SNPs ont été identifiés entre les BAC CIRB286F09 et 22M06. Tous ces indels sont situés dans des zones de répétition du même nucléotide (poly A, poly T, poly C ou poly G).

La séquence du BAC CIRB251D13 est composée de trois contigs représentant au total 130 kb de séquence. L'extrémité R du BAC CIRB251D13 s'ancre à 6 813 pb en amont de l'extrémité R du CIRB286F09, dans un domaine intergénique. L'analyse de la séquence nous a révélé que ce BAC couvre de 11 937 pb l'espace de droite et recouvre totalement le BAC 164H22. Aucun SNP n'a été détecté d'après la comparaison de séquences des BAC CIRB251D13 et CIRB286F09. Ces deux BAC appartiennent donc à l'haplotype cible et





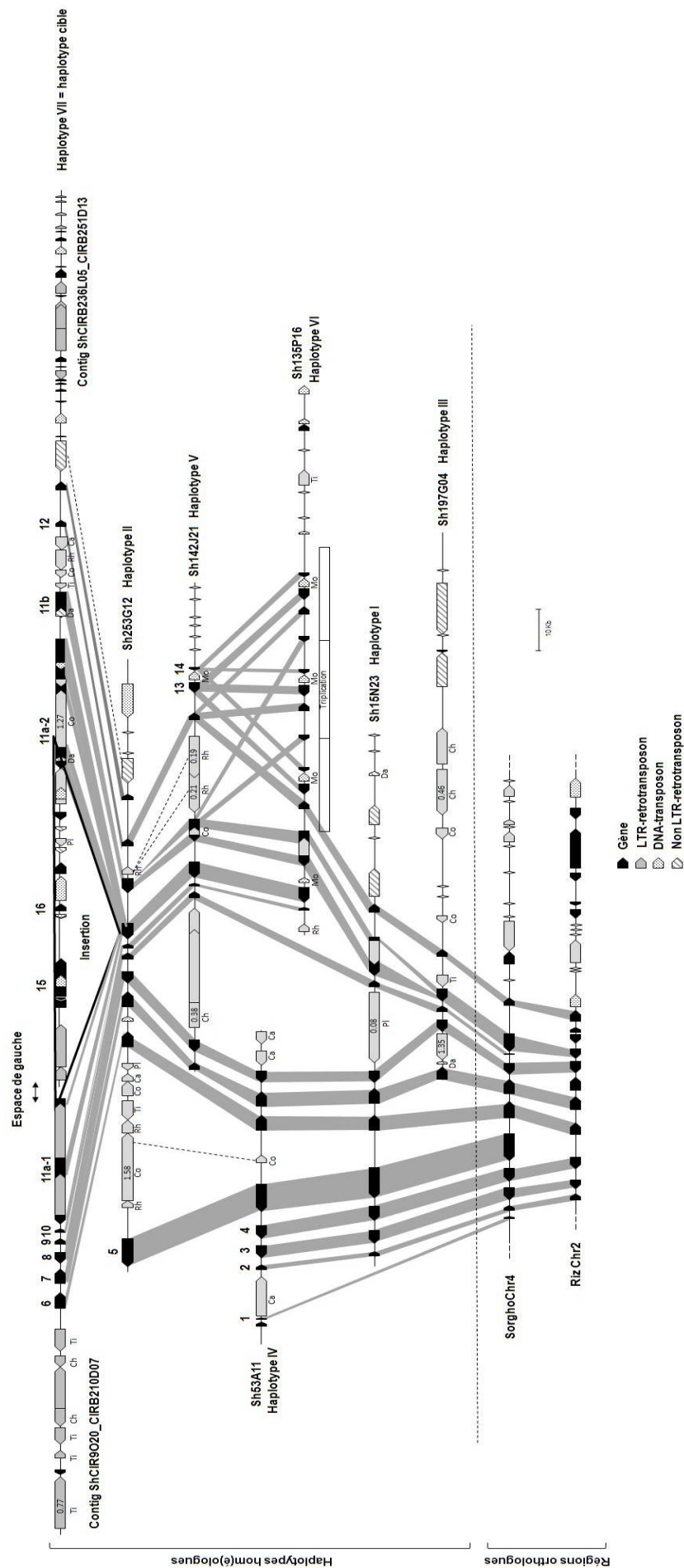
permettent de couvrir en totalité l'espace de droite de la carte physique de l'haplotype cible. Le BAC CIRB286F09 rentre dans l'espace de droite de 29 727 pb et le BAC CIRB251D13 de 11 937 pb. La taille totale de l'espace de droite, préalablement estimée à 30 000 pb d'après les données de synténie entre les haplotypes hom(é)ologues, est donc de 41 664 pb. L'annotation de ces 41 kb a confirmé la présence d'un gène codant pour une endoglucanase de type  $\beta$  1-4 (Figure 2.7). La structure de ce gène (3 exons et 2 introns) correspond à la structure des glucanases présentes sur les autres haplotypes hom(é)ologues (voir chapitre III).

Cette annotation (Figure 2.7) a aussi révélé la présence d'une autre protéine annotée HP qui présente 99% d'identité avec la séquence codant pour une HP située sur le BAC 253G12 mais avec aucune autre séquence des autres haplotypes hom(é)ologues. D'autre part, cette région comprend un rétrotransposon non-LTR de type LINE d'une taille de 8 465 pb et un transposon d'une taille de 2 709 pb conservés avec le BAC 253G12 (98 % et 99 % d'identité) (Figure 2.8).

### **E. Conclusion**

La stratégie basée sur la synténie existant entre les différents haplotypes hom(é)ologues nous a donc permis d'identifier deux nouveaux BAC (CIRB286F09 et CIRB251D13) permettant de combler l'espace de droite de la carte physique de l'haplotype cible (séquence de 41 kb). L'annotation a permis d'identifier une endoglucanase de type  $\beta$  1-4 conservée entre les différents haplotypes hom(é)ologues (Figure 2.8), un gène codant une HP présentant une très forte homologie de séquence avec le BAC 253G12 et deux éléments transposables. La colinéarité entre l'haplotype cible et les autres haplotypes hom(é)ologues est particulièrement bien conservée, notamment avec l'haplotype II (BAC 253G12, 98,7 % d'identité).

Concernant l'espace de gauche, la stratégie basée sur la synténie n'a pas pu être utilisée du fait de la présence d'une insertion sur l'haplotype cible. Une première étape a été de rechercher la région orthologue à l'insertion chez le sorgho mais les résultats ne nous ont pas permis de mettre en évidence une telle région chez le sorgho. La deuxième étape a été de cribler les ressources BAC avec des marqueurs flanquants l'espace de gauche.



**Figure 2.8. Annotation finale des BAC hom(é)ologues de la région de *Bru1* et des BAC des régions orthologues du sorgho et du riz.**  
Les traits pleins grisées représente les relations synténiques existantes entre les différents gènes des BAC. Les traits en pointillés représente les relation synténiques entre les éléments transposables.

Quatre BAC (CIRB9N07, CIRB114G05, CIRB127D08, CIRB210D07) ont été identifiés comme couvrant partiellement le côté gauche de l'espace de gauche et un BAC (CIRB236L05) comme couvrant partiellement le côté droit de l'espace de gauche. Le BAC CIRB236L05 couvre aussi partiellement l'espace de droite. Deux BAC (CIRB210D07 et CIRB236L05) ont été séquencé puis annotés. Le BAC CIRB210D07 a permis d'avancer de 35 kb dans l'espace de gauche. Ce fragment d'ADN comprend le début du gène *S6PDH* 11a-1 qui était tronqué sur le BAC CIR9O20. Le BAC CIRB236L05 n'étant pas complet, nous ne disposons que de 4 kb de séquence couvrant partiellement l'espace de gauche. L'annotation de ces BACs ne nous a pas permis d'identifier de séquence génique.

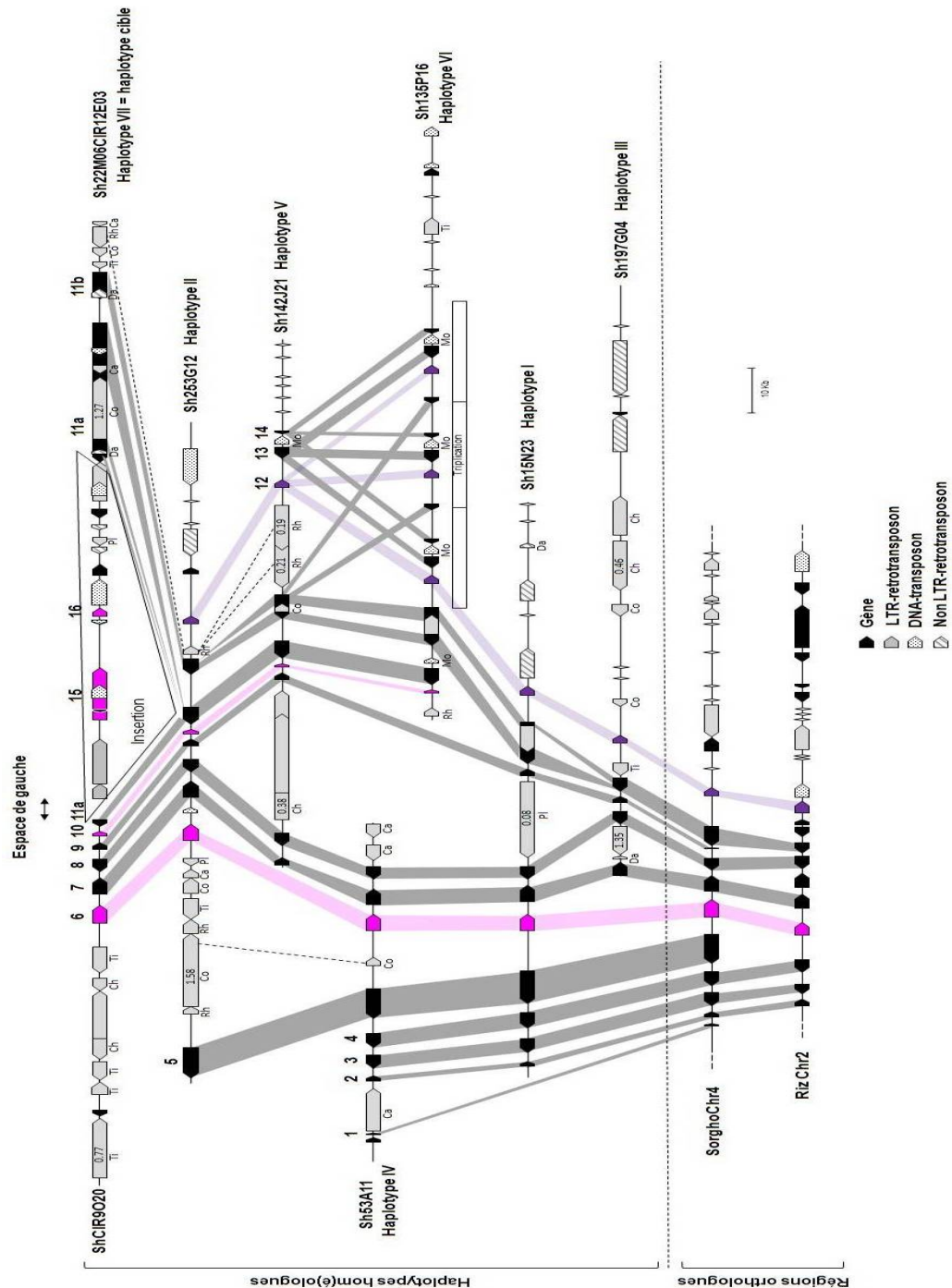
En conclusion, nous avons couvert l'espace de gauche d'environ 40 kb. La carte physique de l'haplotype cible comprend (en allant du côté proximal vers le côté distal) (Figure 2.8) : le contig ShCIR9O20\_CIRB210D07 d'une taille de 127 kb, un espace d'une taille indéterminée et le contig ShCIRB236L05\_CIRB251D13 d'une taille de 300 kb. Nous avons montré, par des analyses PCR, fingerprinting et grâce à l'annotation, que le BAC CIRB210D07 appartient à l'haplotype cible. Mais le nombre de SNP élevé par rapport à la séquence du CIR9O20 indiquerait que ce BAC n'appartient pas à l'haplotype cible. Les résultats obtenus sont assez contradictoires et ne nous permettent pas de valider complètement ce BAC.



---

## **Chapitre 3 : Analyse des gènes candidats présents sur la séquence partielle de l'haplotype cible**

---



**Figure 3.1. Localisation des gènes candidats sur les séquences des BACs de la région de *Brul*.**

Les boîtes noires représentent les gènes et les boîtes grises et rayées les éléments transposables. Les S/T kinases et les glucanases apparaissent en rose et en violet respectivement.

## **Chapitre 3: Analyse des gènes candidats présents sur la séquence partielle de l'haplotype cible**

La mise en place des mécanismes de défense peut se diviser en quatre étapes : reconnaissance du pathogène, régulation, transduction du signal et enfin mise en place des réactions de défense à proprement parlé. Pour chaque étape, certaines familles de gènes ont été décrites chez plusieurs espèces et peuvent être considérées comme des gènes candidats intervenant dans les mécanismes de résistance. Nous avons donc recherché, à partir des annotations automatiques de la séquence partielle de l'haplotype cible et des séquences des haplotypes hom(é)ologues de la région de *Bru1*, la présence de gènes susceptibles d'intervenir dans des mécanismes de résistance.

Nous avons identifié cinq gènes présentant de fortes homologues avec des gènes pouvant intervenir dans les mécanismes de résistance : quatre gènes codant pour une S/T kinase et un gène codant pour une endoglucanase (Figure 3.1).

Afin de déterminer si ces gènes représentent des candidats dans le cas de la résistance *Bru1*, une analyse fine de la structure des cinq gènes candidats et des tests d'expression des gènes validés structuralement ont été réalisés.

### **A.Méthodes**

#### **A.1. Méthodes d'analyses de la structure des gènes**

La séquence des BAC a été annotée automatiquement avec le logiciel de prédiction EuGène (Foissac *et al.*, 2008) qui intègre aussi les informations provenant d'autres logiciels de prédiction tel que FgeneSH ainsi que les analyses de similarités (BLAST : Basic Local Alignment Search Tool) avec les séquences de diverses bases de données (EST, mRNA, séquences protéiques, séquences répétées, etc.), permettant ainsi d'attribuer une fonction aux gènes prédits.

```

S/T_kinase_cas  1  MSPNNDSCEWRDNMDMEAMNP-----LGLLKRITDNFSDRRIIGKGLGVVYK
S/T_kinase_riz  1  -----MNENVP-----YQLLKEITDSFSEERKLGKGGYGTIVYK
Gene_Rpgl      71  PAKEEKKDGDGKDGADKKKEVPVPVPPVYYPFPYPPRPRYEFTVGFINSITDNFSEARIIVRCRHGIVYK
consensus      71  . . . . . * . . . . . * . . . . . * . . . . . * . . . . .

S/T_kinase_cas  52  VWH----EIAVKLFNSTH-KIDEQEDRDCKNLATFKHONI VKLVGFONETDKVTVESNGKFILDCNMHKA
S/T_kinase_riz  35  GEHKNGDKIAVKILHTEG-EDDNKENNEFCNLMMANHTNIIRLVGYCYEIKYTTGLAKCRLALIAITIDRA
Gene_Rpgl      141  GVLDNGQYIAVKLHLMPCIDDEEEKNEFNLMRVRHONIIRLVGYCHHTKQVIVEHNGKHVSARVEERY
consensus      141  . . . . . * . . . . . * . . . . . * . . . . . * . . . . .

S/T_kinase_cas  117  LCVEYLCNGPLCKYIS--VGLNWHERYCIILGICEGLNYLHTRPRSETTYHLNIDLDNMVLLDNMVPKIA
S/T_kinase_riz  104  LCFEYMPFGSLNLYLSDDEGLDWPIRYKIKGTCEGLKHLHVEMPPYIYHLIDKPANILLDNMVPKIA
Gene_Rpgl      211  LCSEYLECGSLDKHLSNEPCALAWYTCYKIKGICDGLLCLHKGFBPIIWHLEIKPTKILLDKDMMPKIG
consensus      211  * . * . * . * . . . . * . . . . . * . . . . . * . . . . . * . . . . .

S/T_kinase_cas  184  DFGLSTLEHKES----ISFEGARGYMPPEYIRDQETSKFDFSLDVIIKIMVGHMPLYLVINGMGSTEV
S/T_kinase_riz  174  DFGLSRFITEEKTMIITKNVGTLPYMPREYIERKEISNKLDFSLGAVMLDIAGPTGRFRSTKMPSLEF
Gene_Rpgl      281  GFGESRLSDSIETSSSTSEVRGTSVYMPPEYISKR-ITSSNVFSLGVIIIQIMAGKESYTKCVDIPPEEF
consensus      281  . * . * . . . . . * . . . . . * . . . . . * . . . . . * . . . . .

S/T_kinase_cas  250  IEFVHNKWRRLRMVPESISTEVOQQVKGCDGVLR-----
S/T_kinase_riz  244  TNQVQGWKNRIQTRRNGPILLEAYQQVKTCTETALKCLEIDRENRPNIETINQINEKEDIIGELPIDH
Gene_Rpgl      350  TERVYGFVNRMPG---TVSKHTSEVKTCTETALKCVENRVNRPTINEIICRMKDIDIVECSSIGDL
consensus      351  . . . * . . . . . . . . . . * . . . . . . . . . . . . . . . .

```

**Figure 3.2. Alignement des séquences en acides aminés des protéines S/T kinases de canne à sucre, de riz et d'orge.**

I à XI : sous-domaines caractéristiques des kinases

cas = canne à sucre



L'expertise manuelle des gènes prédits par EuGène est réalisée en visualisant les annotations automatiques dans l'éditeur d'annotation Artemis (Rutherford *et al.*, 2000). La structure des gènes est vérifiée quant à la pertinence de la délimitation des segments correspondant aux introns et aux exons des gènes prédits (ATG, codons stop, sites d'épissage). La nature de ces gènes est vérifiée en comparant leurs séquences avec des bases de données nucléiques (base de données nr/nt : « Nucleotide collection », algorithme blastn) et protéiques (base de données Swissprot, algorithme blastp). Une comparaison de séquences sur la base de données d'EST de canne à sucre ([http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=s\\_officinarum](http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=s_officinarum)) est également réalisée avec l'algorithme tblastx afin de rechercher d'éventuels ADNc correspondants. L'analyse de la structure des kinases est réalisée en effectuant un alignement multiple de leur séquence protéique (ClustalW) avec les séquences de différentes sérine/thréonine kinases fonctionnelles de différentes espèces (Figure 3.2). Cet alignement permet de déterminer le taux de conservation des sous-domaines de ces kinases.

## A.2. Extraction des ARN

Les ARN totaux sont extraits à partir de tissus végétaux en utilisant le kit TRI-Reagent® (Sigma-Aldrich, St Louis, USA). Cette technique d'extraction permet d'obtenir, à partir de 100 mg de matériel frais, des ARN purs, de bonne qualité et en forte concentration. Le matériel végétal est broyé dans l'azote liquide puis homogénéisé avec 500 µL de solution TRI-Reagent®. Le TRI-Reagent® est une solution acide qui permet de solubiliser l'ARN total et ainsi de le séparer de l'ADN et des protéines. Après une étape d'extraction au chloroforme, les ARN sont précipités à l'isopropanol froid. Les culots d'ARN obtenus sont lavés dans 1 volume d'éthanol à 70% puis solubilisés dans de l'eau *RNase free*. Les ARN totaux obtenus sont ensuite traités à la DNase (RQ1 RNase-Free DNase de Promega) pour éliminer tout l'ADN génomique restant. Les conditions utilisées sont celles proposées par le fabricant. La qualité des ARN extraits est analysée sur gel d'agarose. Leur quantité et leur pureté sont estimées par analyse spectrophotométrique à 260 et 280 nm (1 unité de DO<sub>260</sub> = 40 µg d'ARN; les ARN sont de bonne qualité si le rapport DO<sub>260</sub> / DO<sub>280</sub> est compris entre 1,6 et 2).



### A.3. Obtention des ADNc

Les ADNc sont obtenus à partir de 2 µg d'ARN totaux en utilisant le kit d'Invitrogen SuperScript® III Reverse Transcriptase. Ce kit permet de synthétiser le premier brin d'ADN complémentaire à partir d'une extraction d'ARN totaux. Ce kit a été optimisé pour réduire l'activité RnaseH et fournit une plus grande stabilité thermique. Les ARN totaux sont mis en incubation à 65°C pendant 5 minutes avec un mélange d'1 µl d'oligo (dT)<sub>20</sub> à 50 µM, 1 µl dNTP 10 µM et qsp H<sub>2</sub>O pour compléter à un volume finale de 10 µl. Cette étape permet la dénaturation des ARN totaux. Ce mélange est ensuite incubé à 50°C pendant 50 minutes avec un mélange de 2 µL de tampon de réaction 10X RT Buffer, 4 µL de MgCl<sub>2</sub> 25mM, 2 µl de DTT 0,1M, 1 µl RNaseOUT (40 U/µl) et 1 µl SuperScript® III RT (200 U/µl). La réaction est arrêtée en plaçant l'échantillon dans un bain à 85 °C pendant 5 min. La qualité des ADNc est ensuite vérifiée par amplification PCR en prenant un gène de référence comme témoin.

### A.4. Amplification par PCR

Les réactions PCR sont réalisées dans un volume final de 50 µl avec les composants suivants : 2,5 µL d'ADN génomique à 10 ng.µL<sup>-1</sup> ou 2,5 µL d'ADNc, 5 µL de tampon de réaction 10X (Eurobio), 0,5 µL de MgCl<sub>2</sub> à 25 mM, 4 µL de dNTP à 2 mM, 2,5 µL d'amorce sens et antisens à 10 µM, et 1 unité de *Taq* DNA polymérase (Eurobio). Le programme d'amplification PCR commence par une étape de dénaturation de l'ADN à 94°C pendant 4 min, puis 35 cycles à 94°C pendant 30s, 55 °C pendant 30s, 72°C pendant 45s et une extension finale à 72°C pendant 8 min. Les produits PCR sont visualisés sur un gel d'agarose à 2 %.

## **B. Endoglucanase de la région de *Bru1* (gène 12)**

### **B.1. Implication des endoglucanases dans les mécanismes de résistance**



Les endoglucanases sont des enzymes appartenant au groupe des glycosides hydrolases (EC3.2.1) comprenant plus de 150 enzymes différentes. Ces enzymes interviennent dans divers processus métaboliques comme le métabolisme des parois cellulaires, dans les mécanismes de défense et dans la voie de biosynthèse et d'hydrolyse de l'amidon (Minic, 2008).

Il existe deux types d'endoglucanases qui se différencient par le substrat qu'elles hydrolysent :

- $\beta$ -1,3- endoglucanase : hydrolyse le  $\beta$ -1,3-glucane. Ces enzymes interviennent principalement dans les interactions plantes/ microorganismes (Münch-Garthoff *et al.*, 1997; Lepoivre, 2003; van Loon *et al.*, 2006). Leur action est de deux types : (1) une dégradation des parois fongiques entraînant la réduction de la croissance de l'agent pathogène et (2) une libération de molécules signal du parasite (éliciteur) permettant d'amplifier la réponse de défense (Lepoivre, 2003).

- $\beta$ -1,4- endoglucanase : hydrolyse le  $\beta$ -1,4-glucane. Ces enzymes interviendraient dans des mécanismes cellulaires de développement de l'organisme (Minic, 2008). Sa principale fonction serait dans l'élongation des parois des cellules en participant à la formation de la cellulose, principal constituant des parois des cellules végétales (Levy *et al.*, 2002). De part leur fonction, ces gènes sont très conservés entre différentes espèces (Minic, 2008). Ce type de gène n'a pas été montré comme intervenant dans les mécanismes de résistance.

La différence de structure de ces deux enzymes dans une même espèce peut être très difficile à caractériser (Henrissat *et al.*, 2001). Une étude portant sur deux endoglucanases d'orge, une  $\beta$ -1,3- endoglucanase (fonction dans les mécanismes de défense) et une  $\beta$ -1,4- endoglucanase (fonction dans la synthèse des parois cellulaires), a montré que ces deux protéines avaient une structure secondaire très proche (Varghese *et al.*, 1994). Les différences se situent sur quelques acides aminés dans les sites catalytiques de l'enzyme qui entraînent une spécificité du substrat différente.

**Figure 3.3. Alignement des séquences en acides aminés de l'endoglucanase présente sur l'haplotype cible de la région de *Bru1* de canne à sucre et des  $\beta$ -1,4- endoglucanases et des  $\beta$ -1,3- endoglucanases présentes chez le sorgho, le riz, le peuplier et l'arabette.**  
cas = canne à sucre

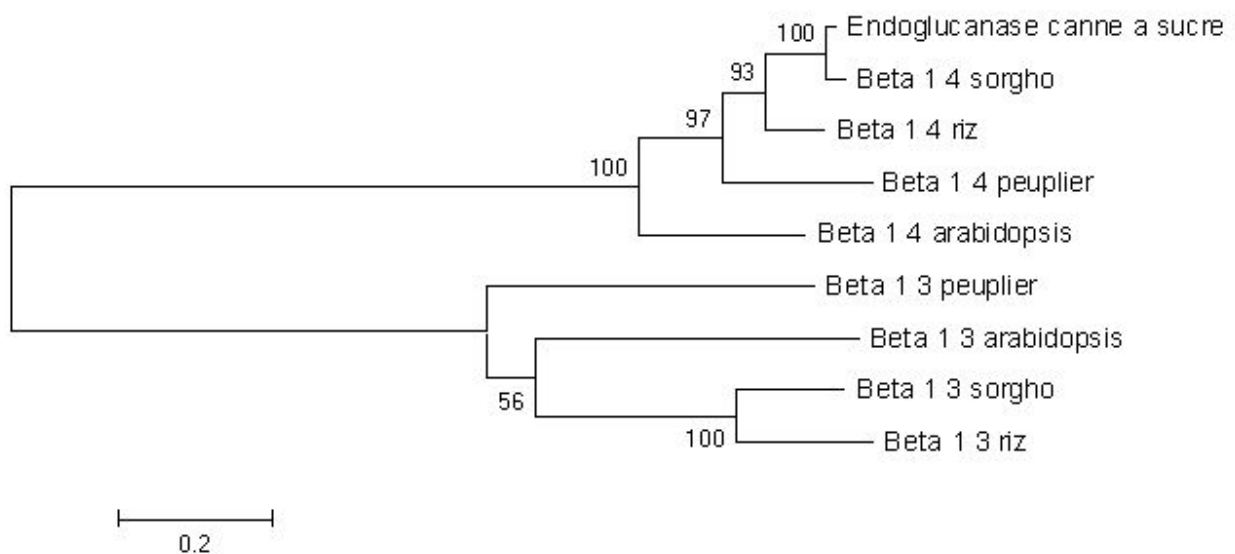
## **B.2. Analyse structurale de l'endoglucanase de la région de *Bru1***

Le gène codant pour l'endoglucanase est présent sur six haplotypes hom(é)ologues (Figure 3.1). Ce gène comprend 3 exons et 2 introns. Sa taille moyenne génomique sur les différents haplotypes hom(é)ologues est de 1,8 kb pour un transcrit moyen prédit de 1,5 kb. Afin de déterminer si cette glucanase appartient au groupe des  $\beta$ -1,3- ou des  $\beta$ -1,4-endoglucanases, nous avons comparé sa séquence nucléotidique avec les endoglucanases présentes dans les bases de données. Les alignements de séquences montrent une identité supérieure à 60% avec des  $\beta$ -1,4-endoglucanases, l'homologie la plus forte correspondant à une  $\beta$ -1,4-endoglucanase du sorgho orthologue à la région de *Bru1* (96 % d'identité, Figure 3.1). Une homologie de séquence a également été retrouvée (72 % d'identité) avec la  $\beta$ -1,4-endoglucanase de riz présente dans la région orthologue du chromosome 2 (Figure 3.1). Pour compléter cette analyse, nous avons réalisé un alignement des séquences en acides aminés des  $\beta$ -1,4-endoglucanases et des  $\beta$ -1,3-endoglucanases de sorgho et de riz et de deux dicotylédones, *Arabidopsis thaliana* et *Populus trichocarpa*, les  $\beta$ -1,4-endoglucanases constituant une famille de gènes très conservée entre espèces (Figure 3.3). L'endoglucanase au niveau de la région de *Bru1* présente une très forte homologie de séquence avec les endoglucanases de type  $\beta$ -1,4 des deux Poacées mais aussi avec les endoglucanases de type  $\beta$ -1,4 des deux espèces des dicotylédones. Ce résultat est conforté avec l'arbre phylogénétique dérivé de l'alignement des séquences en acides aminés (Figure 3.4). L'ensemble de ces résultats confirme que la glucanase codée par le gène 12 appartient au groupe des  $\beta$ -1,4-endoglucanases et n'interviendrait donc pas dans les mécanismes de résistance.

## **C. Kinases présentes sur l'haplotype cible (gènes 6, 10, 15, 16)**

### **C.1. Implication des kinases dans les mécanismes de résistance**

Les protéines kinases constituent l'une des plus grandes familles de protéines dans les génomes eucaryotes. Elles catalysent les réactions de phosphorylation par l'ajout d'un ion phosphate à un substrat, entraînant ainsi une modification des propriétés fonctionnelles du



**Figure 3.4. Arbre phylogénique obtenu à partir de l'alignement des séquences en acides aminés des endoglucanases de canne à sucre, sorgho, riz et peuplier (cf figure 3.3).**

L'arbre a été construit en utilisant la méthode Neighbor-joining implémentée dans le logiciel MEGA4. La robustesse de la topologie de l'arbre est estimée avec un bootstrap de 1000 réplicats. La barre d'échelle représente la distance génétique relative (nombre de substitutions par nucléotide).

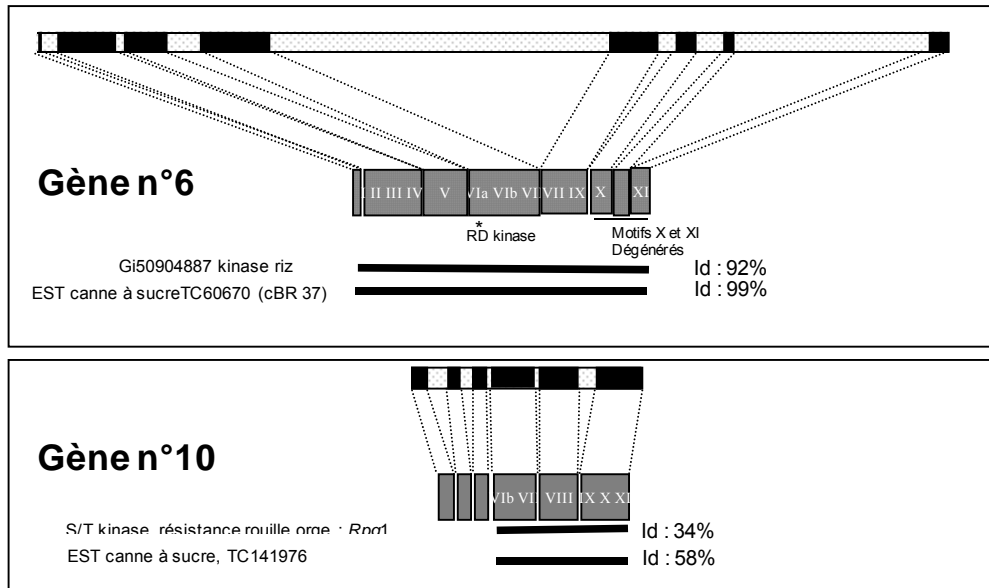


substrat. Elles sont au centre de réseaux cellulaires de signalisation qui coordonnent diverses activités telles que le métabolisme de réponse au stress biotique ou abiotique, la transcription, la traduction, la réplication de l'ADN, le contrôle du cycle cellulaire, le développement d'organes... La structure de ces protéines est composée de sous-domaines plutôt bien conservés entre espèces. Les kinases contiennent 12 sous-domaines : I, II, III, IV, V, VIa, VIb, VII, VIII, IX, X et XI, les domaines X et XI étant les moins conservés (Hanks & Hunter, 1995) (Figure 3.2) Les domaines I à IV sont impliqués dans la fixation et l'orientation de l'ATP et les domaines VIa à XI sont responsables de la fixation du substrat peptidique. De nombreuses études ont montré l'implication de cette classe de protéines dans les mécanismes de résistance, en tant que récepteurs dans l'étape de reconnaissance du pathogène. Les récepteurs kinases ont été séparés en deux classes, sur la base de la présence ou non de deux acides aminés (Arginine, R, Aspartate, D) dans le sous-domaine VI-a (Dardick & Ronald, 2006). Ces deux classes sont appelées respectivement « Kinase RD » et « Kinase non RD ». Dardick & Ronald (2006) ont montré que les récepteurs kinases interagissant avec un agent pathogène sont de type « Kinase non RD ». Deux grands types de protéines de résistance à domaine kinase peuvent être identifiés en fonction de la localisation cellulaire de leurs produits : les protéines membranaires et les protéines cytoplasmiques.

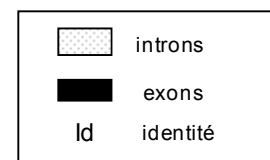
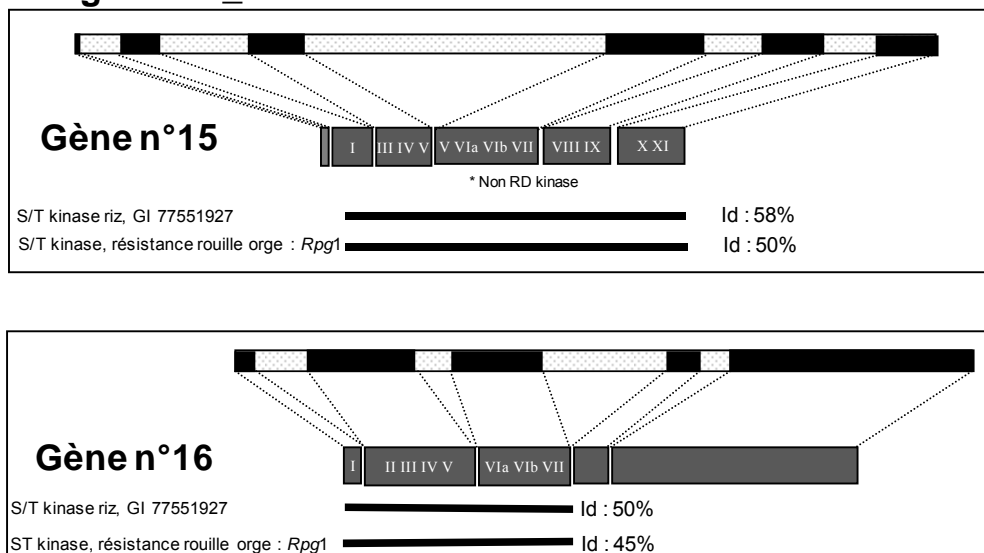
- Protéines membranaires : en plus du domaine kinase intracellulaire, toutes ces protéines possèdent un domaine transmembranaire qui ancre la protéine dans la membrane cytoplasmique. Elles possèdent un domaine extracellulaire, souvent de type LRR (Leucine Rich Repeat) ou non. Par exemple, le gène *Xa21* impliqué dans l'interaction riz / *Xanthomonas oryzae* code pour une protéine de type LRR-kinase (Song *et al.*, 1995).

- Protéines cytoplasmiques : ces protéines ne présentent pas de domaine extracellulaire. Le premier gène de résistance de ce type à avoir été cloné est le gène *Pto* impliqué dans l'interaction tomate / *Pseudomonas syringae*. Ce gène code pour une kinase intracellulaire de type kinase RD (Martin *et al.*, 1993). Un autre gène codant pour une kinase cytoplasmique a été cloné. Il s'agit du gène *Rpg1* dans l'interaction orge / *Puccinia graminis* (Brueggeman *et al.*, 2002). La structure de la protéine RPG1 montre deux kinases intracellulaires fonctionnant en tandem, dont l'une est de type kinase non RD et l'autre de type ACF (Alternative Catalytic Function) (Dardick & Ronald, 2006).

## BAC Cir9020



## Contig 22M06\_CIR12E03



**Figure 3.5. Séquences présentes sur les BACs de la région cible de *Bru1* chez la canne à sucre présentant des homologies de séquences avec des sérine/thréonine kinases.**  
Les zones d'homologie protéique sont représentées par des traits horizontaux.

Il existe aussi des kinases pouvant intervenir dans la transduction du signal de défense : les MAPKinases (MAPK). Elles sont décrites comme intervenant très tôt dans la cascade des réactions de défense (Zhang & Klessig, 2001).

## C.2. Analyse structurale des kinases présentes sur l'haplotype cible

Les quatre kinases révélées lors de l'annotation automatique présentent une homologie de séquence avec des S/T kinases. L'alignement des séquences en acides aminés de ces kinases avec les différentes S/T kinases fonctionnelles correspondant au gène *Pto* chez la tomate et au gène *Rpg1* chez l'orge (Martin *et al.*, 1993; Brueggeman *et al.*, 2002) a permis de repérer les sous-domaines communs aux quatre protéines. L'annotation manuelle est schématisée sur la figure 3.5.

### C.2.1 S/T kinase codée par le gène 6 (BAC CIR9O20)

La S/T kinase codée par le gène 6 est présente sur l'haplotype cible au niveau du BAC CIR9O20 mais aussi sur trois des six autres haplotypes hom(é)ologues de la région de *Bru1*. Nous avons également retrouvé les gènes orthologues chez le sorgho (chromosome 4) et le riz (chromosome 2) (Figure 3.1). Le gène 6 est composé de huit exons et de sept introns. Sa taille génomique est de 4,3 kb et la taille du CDS est de 882 pb. La protéine codée par le gène présente 92 % d'identité avec une sS/T kinase impliquée dans le contrôle de la division cellulaire chez *O. sativa* : P34 cell division control 2 (gi50904887). L'étude approfondie de sa structure nous a indiqué que cette kinase possède les 12 sous-domaines I à XI et présente un motif RD dans le sous-domaine VIb. Les kinases présentant ce motif sont généralement impliquées dans la croissance cellulaire et le développement (Dardick & Ronald, 2006). D'autre part, ce gène présente 99 % d'identité nucléique avec une séquence EST de canne à sucre (TC60670) et avec le cDNA (SCRLST3163H06) qui correspond au marqueur RFLP cBR37 cartographié à 0,28 cM du gène *Bru1*. Ces résultats nous indiquent que ce gène semble être fonctionnel mais qu'il n'est pas un bon candidat pour la résistance à la rouille.

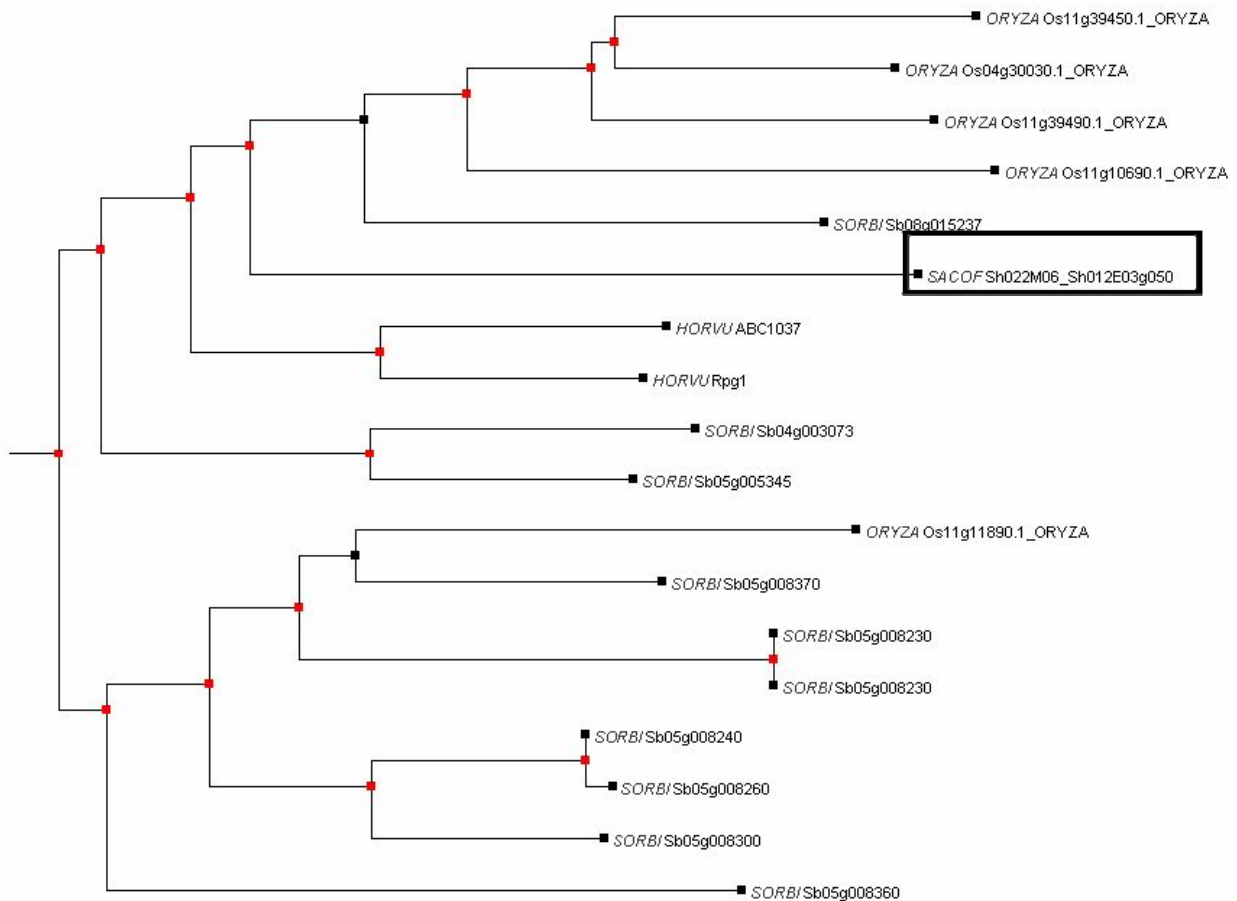


### C.2.2 S/T kinase codée par le gène 10 (BAC CIR9O20)

La S/T kinase codée par le gène 10 est présente sur quatre des sept haplotypes hom(é)ologues de la région de *Bru1*. Sur l'haplotype cible, elle est située sur le BAC CIR9O20. Cette kinase n'a pas été retrouvée dans les régions du sorgho et du riz orthologues à la région de *Bru1* (Figure 3.1). Le gène 10 est constitué de six exons et de cinq introns. Sa taille génomique est de 1,1 kb et la taille du CDS prédit est de 534 pb. L'alignement de sa séquence en acides aminés avec les S/T kinases présentes dans les bases de données a permis de démontrer que seule une partie de la protéine (exons 4 à 6) présente une homologie avec les sous-domaines VIb à XI des S/T kinases. Les sous-domaines I à VIa ne sont pas présents. La protéine présente une homologie de séquence avec une S/T kinase impliquée dans les mécanismes de résistance : 34 % d'identité protéique avec le gène *Rpg1* au niveau des exons 4 à 6 et une homologie partielle (58 % d'identité) avec une EST de canne à sucre (TC141976) au niveau des exons 4 à 6. Les trois premiers exons n'ont pas présenté d'homologie avec des domaines protéiques connus. Cette S/T kinase n'est pas complète (absence des 12 sous-domaines) et semble être un pseudogène. De ce fait, elle ne nous paraît pas être un bon candidat.

### C.2.3 S/T kinase codée par le gène 15 (contigs 22M6\_CIR12E03)

La S/T kinase codée par le gène 15 se situe uniquement sur le contig 22M06\_CIR12E03 dans l'insertion spécifique à l'haplotype porteur du gène *Bru1* (Figure 3.1). Ce gène possède six exons et cinq introns. Sa taille génomique est de 12,8 kb et la taille du CDS prédit est de 858 pb. Sa grande taille génomique s'explique par la présence d'un intron de plus de 11 kb entre les exons 3 et 4. A l'intérieur de cet intron, nous retrouvons un fragment de gène correspondant à une enzyme malique NADP-dépendant et un transposon. Cette structure de gène a déjà été décrite par Brueggeman *et al*, (2006). Les auteurs ont identifié des gènes appartenant à la famille du gène *Rpg1* chez l'orge dont un qui présentait dans sa structure un très grand intron entre l'exon 3 et l'exon 4 (12,5 kb). L'alignement des séquences en acides aminés montre une homologie avec une S/T kinase de riz (58 % d'identité protéique) et avec la protéine RPG1 de l'orge (50 % d'identité protéique). Ces



**Figure 3.6.** Arbre phylogénétique obtenu à partir de l'alignement des séquences en acides aminés des protéines orthologues à la protéine codée par le gène 15.

Cet alignement a été réalisé en utilisant l'outil i-GOST de la base de données GreenphylDB (adresse internet).

SORBI = *Sorghum bicolor*, SACOF = *Saccharum officinarum*, HORVU = *Hordeum vulgare*, ORYZA = *Oryza sativa*.

alignements nous ont permis d'identifier les 12 sous-domaines caractérisant les kinases, les domaines X et XI étant dégénérés (Figure 3.2). Le domaine VIb ne présente pas de motif RD. Cette kinase appartient donc à la famille des kinases non-RD impliquées dans les mécanismes de résistance (Dardick & Ronald, 2006). Nous avons blasté la séquence CDS du gène 15 sur les bases de données EST disponibles, notamment pour bien délimiter la fin du gène. Mais aucune EST n'a été identifiée. Afin de rechercher des séquences protéiques orthologues, nous avons utilisé l'outil GOST (GreenPhyl Orthologs Search Tool) de la plateforme GreenPhylDB développée au CIRAD (<http://greenphyl.cirad.fr>) : cette base de données est conçue pour faciliter les études de génomique comparative et fonctionnelle chez les plantes (Conte *et al.*, 2008; Rouard *et al.*, 2010). La principale fonction de GreenPhylDB est d'attribuer à un jeu de séquences protéiques une famille de protéines définies dans la base de données puis par la suite de créer des groupes afin d'identifier le plus proche orthologue en utilisant une approche phylogénomique. La séquence de la protéine codée par le gène 15 a été comparée à des séquences protéiques de riz et d'*Arabidopsis*. Le plus proche orthologue de cette protéine est situé sur le chromosome 11 du riz (58 % d'identité protéique - Os11g39450.1). Cette séquence correspond à une protéine de type récepteur kinase. La séquence protéique du gène Os11g39450.1 a été blastée sur NCBI (BLASTP, <http://www.ncbi.nlm.nih.gov/blast/>) pour rechercher des orthologues (e-value supérieure à 1E-10). Deux séquences ont ainsi été identifiées chez l'orge, dont la séquence du gène *RpgI*, 10 séquences ont été identifiées chez le sorgho, 4 autres séquences ont aussi pu être identifiées chez le riz. Un arbre phylogénétique a été construit à partir de ces séquences avec l'outil i-GOST de GreenphylDB (Figure 3.6). Le plus proche orthologue du gène 15 est un gène présent sur le chromosome 8 du sorgho. Le gène du sorgho a une taille de 16 kb avec deux grands introns. Il code pour une protéine ayant une fonction S/T kinase.

Afin d'approfondir cette analyse, nous avons recherché autour du gène 15 s'il existait d'autres domaines protéiques composant les protéines de résistance comme des domaines LRR. Dans un premier temps nous avons comparé une zone de 10 kb en amont de l'exon 1 du gène 15 avec les bases de données nucléotidiques et protéiques existantes. Aucune séquence ne correspond à une structure de gène connu. Dans un second temps, nous nous sommes intéressés à une zone de 14 kb en aval de l'exon 6. La séquence obtenue pour cette partie comporte trop de codons stop et est donc inexploitable pour un BLAST. Nous avons analysé cette séquence à la main pour vérifier la présence de motifs caractérisant les domaines LRR





des protéines membranaires. Mais aucun motif n'a été mis en évidence.

Ainsi, seul le gène 15 code pour une protéine fonctionnelle potentiellement impliquée dans la résistance dans cette zone de 37 kb, protéine avec un domaine S/T kinase comprenant les douze sous-domaines fonctionnels. Les comparaisons de séquences nous ont indiquées que le gène 15 présente une homologie avec un gène connu comme intervenant dans la résistance à la rouille chez l'orge et de nombreuses kinases. De plus, le sous-domaine VIb ne présente pas de motif RD. Ce gène est donc un bon gène candidat.

#### **C.2.4 S/T kinase codée par le gène 16 (contigs 22M6\_CIR12E03)**

Le gène 16 se situe, comme le gène 15, uniquement sur le contig 22M06\_CIR12E03 dans l'insertion spécifique de l'haplotype cible (Figure 3.1). Le gène 16 est constitué de cinq exons et de quatre introns. Sa taille génomique est de 1,9 kb et la taille du CDS est de 1,2 kb. La kinase codée par ce gène présente une homologie de séquence avec la S/T kinase du riz homologue au gène 15 (50 % d'identité protéique) et présente 45 % d'identité protéique avec la protéine RPG1. La séquence nucléotidique s'aligne partiellement avec une séquence EST de canne à sucre (TC 51269), au niveau des trois premiers exons (77 % d'identité nucléique). Les comparaisons de séquences protéique et nucléique avec RPG1/*Rpg1* nous a permis de mettre en évidence que cette kinase ne présente que les sous-domaines I à VII, correspondant aux exons 1 à 3. Le domaine VIb ne présente pas de motif RD et appartient donc à la famille des kinases non RD. Cette kinase présente aussi un décalage de phase de lecture à partir du quatrième exon. Le BAC 22M06 a été séquencé par la méthode Sanger avec un taux de recouvrement de 12X, l'analyse manuelle de fichiers de sortie du séquençage nous a permis de confirmer la présence du décalage de phase de lecture. La séquence protéique codée par les exons 4 et 5 ne correspond pas à un domaine protéique connu. La protéine prédite ne semble donc pas fonctionnelle ; le gène 16 n'est donc pas un bon candidat.



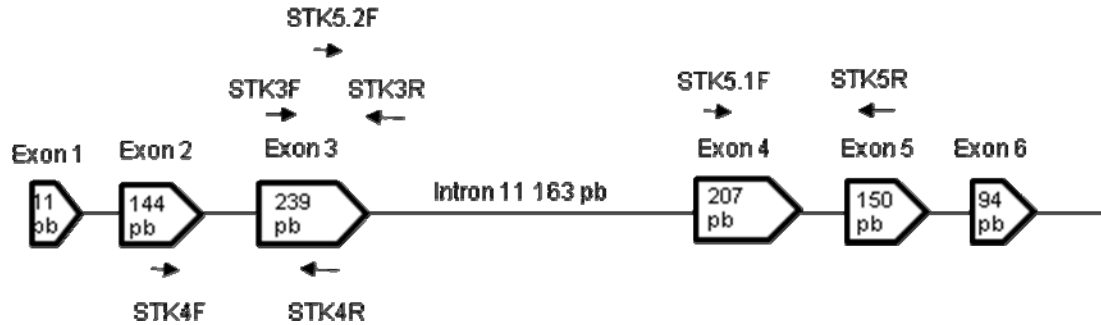
### C.3. Analyse fonctionnelle de la kinase candidate codée par le gène 15

L'analyse structurale des gènes candidats a permis d'identifier un gène candidat qui semble fonctionnel : le gène 15 sur le contig 22M06\_CIR12E03 dans l'insertion spécifique à l'haplotype cible qui code pour une S/T kinase (Figure 3.1). Les comparaisons de séquences nous ont indiqué que cette kinase présentait une forte homologie avec des protéines codées par un gène de résistance à la rouille chez l'orge, le gène *Rpg1* (Brueggeman *et al.*, 2002) et par un gène de riz annoté comme un récepteur kinase (homologie de séquence protéique avec la protéine RPG1). Sur ces deux gènes, seul le gène *Rpg1* a pu être caractérisé et étudié au niveau de son expression.

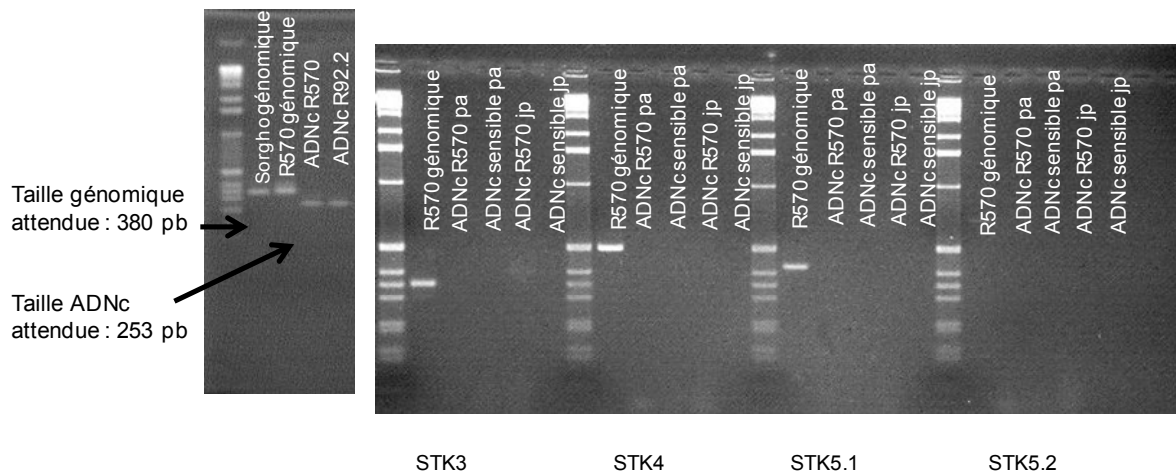
Le niveau d'expression du gène *Rpg1* d'une variété d'orge résistante a été évalué sous condition normale (sans infection) et sous condition d'infection (inoculation du pathogène) (Rostoks *et al.*, 2004). La cinétique d'expression est comparée à un gène intervenant dans les mécanismes de résistance, le gène codant pour la  $\beta$ -1,3-endoglucanase. *Rpg1* est détectable en condition normale dans une plante résistante. Le niveau d'expression du gène *Rpg1* sous condition d'inoculation est identique à celui observé en condition normale, tandis que le gène codant pour la  $\beta$ -1,3-endoglucanase voit son niveau d'expression augmenter fortement sous condition d'inoculation.

Une variété d'orge sensible à la rouille a été transformée par *Agrobacterium tumefaciens* contenant l'ADN génomique du gène *Rpg1* (Horvath *et al.*, 2003). Les transformants obtenus ont été évalués pour quantifier le niveau d'expression du gène sous condition normale. Les niveaux d'expression ont ensuite été normalisés par rapport au niveau d'expression du cultivar sensible (expression à 0 %). Les résultats ont montré que (i) le niveau d'expression de *Rpg1* dans un cultivar résistant est estimé à 100 % par rapport au cultivar sensible (ii) que le niveau d'expression est 25 à 500 % plus important que celui du cultivar sensible suivant le nombre de copies du gène *Rpg1* insérées dans les transformants.

D'après ces études, le gène *Rpg1* est exprimé constitutivement chez l'orge en condition normale (sans inoculation) et est détectable par amplification PCR. Par contre, Brueggeman *et al.*, (2002) ont criblé une banque d'ADNc de 500 000 clones et n'ont détecté que trois clones correspondant à *Rpg1*. Il semble donc que ce gène soit exprimé à un faible niveau.



**Figure 3.7. Structure du gène 15 codant la kinase candidate et localisation des couples d'amorces définis pour les tests d'expression.**



**Figure 3.8. Tests d'expression des ADNc du gène 15.**  
pa = plante âgée, jp = jeune plant

### C.3.1. Test d'expression du gène 15

Pour tester l'expression du gène 15 codant pour la kinase candidate, deux choix s'offraient à nous : (1) test d'expression différentielle sans inoculation du champignon ou (2) test d'expression différentielle avec inoculation du champignon. La synthèse bibliographique effectuée sur le gène *RpgI* de l'orge nous a révélé que ce gène était exprimé constitutivement dans les cellules de l'orge. De plus, cette technique est plus facile à mettre en œuvre que des tests sous condition d'inoculation. Les tests d'expression du gène 15 chez un cultivar de canne à sucre résistant (R570), avec comme témoin un cultivar de canne à sucre sensible (R926246), ont donc été réalisés dans des conditions sans inoculation du pathogène.

Dans un premier temps, nous avons réalisé des extractions d'ARN totaux sur les deux cultivars. Ces extractions d'ARN ont été effectuées à deux stades de développement différents des plantes : stade jeune plantule (plant de canne sorti de culture *in vitro*) et stade plante âgée (plant de canne cultivé en serre depuis plus d'un an). En effet, le niveau de résistance peut-être différent entre une plante juvénile et une plante plus âgée (Develey-Riviere & Galiana, 2007). Les ADNc ont été synthétisés par transcription inverse à partir des ARN totaux. Les ADNc ainsi obtenus ont été testés par amplification PCR en utilisant un couple d'amorces ciblant un gène que l'on sait être exprimé constitutivement dans les cellules : le gène de l'actine. Les amorces ciblant le gène de l'actine ont été définies sur la base d'une comparaison de séquences entre une actine de sorgho et des EST de canne à sucre. L'amplification PCR nous a permis de valider la qualité de nos ADNc. En effet, comme le montre la figure 3.8, nous avons observé que les ADNc provenant des cultivars résistants et sensibles ont un amplificat de 253 pb (taille de l'amplificat de l'ADNc actine attendue) au lieu des 380 pb que l'on observe sur les ADN génomiques (taille de l'amplificat de l'ADN génomique attendue). Dans un second temps, à partir de la structure de la kinase candidate codée par le gène 15, nous avons défini quatre couples d'amorces ciblant les exons du gène candidat. N'étant pas certains de la délimitation des exons, nous avons dessiné les amorces à l'intérieur des exons plutôt qu'en bordure (Figure 3.7). Ces couples d'amorces ont été testés sur l'ADN génomique de R570, l'ADNc de R570 de plantes âgées et jeunes plants et sur l'ADNc du cultivar sensible de plantes âgées et jeunes plants. Les résultats d'amplification sont présentés dans la figure 3.8. Le couple témoin STK3 n'a bien amplifié que l'ADN génomique de R570.



Par contre, pour tous les autres couples d'amorces, aucune amplification n'a été observée sur les ADNc du cultivar résistant et du cultivar sensible.

### C.3.2. Conclusion

Les résultats obtenus par PCR nous ont indiqué qu'aucun des couples ciblant la kinase candidate n'amplifie les ADNc de R570. Trois hypothèses peuvent être formulées :

- le gène de la kinase candidate n'est pas exprimé dans le cultivar résistant R570 et ne serait donc pas le gène de résistance *Bru1* recherché.
- l'expression de ce gène est très faible et nous n'arrivons pas à la détecter par amplification PCR.
- le gène 15 n'est exprimé qu'en condition d'infection. Par exemple, dans le cas du gène de résistance *Xa21*, le niveau d'expression n'est détectable par amplification PCR qu'à partir du cinquième jour après inoculation par le pathogène (Yoshimura *et al.*, 1998).





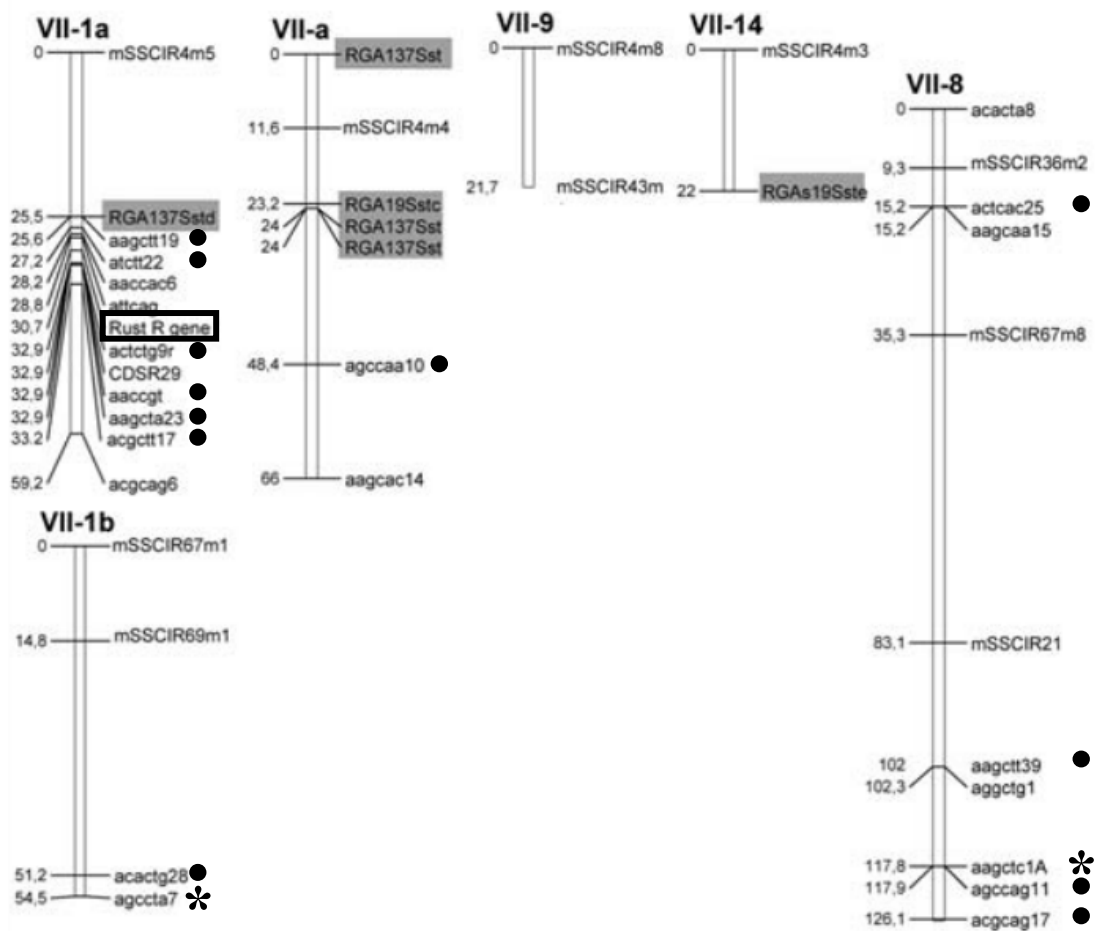
---

---

## **Chapitre 4 : Origine de l'insertion contenant le gène *Bru1***

---

---



**Figure 4.1. Origine des marqueurs moléculaires sur le groupe d'homologie VII du cultivar R570** (Hoarau *et al.*, 2001; Rossi *et al.*, 2003).

Les ronds noirs indiquent les marqueurs dont l'origine putative est *S. officinarum* et les astérisques indiquent ceux dont l'origine est *S. spontaneum*. Le locus de résistance à la rouille brune *Bru1* sur le groupe d'homologie VII-1a est encadré. Les distances génétiques sont exprimées en centiMorgans.

## Chapitre 4 : Origine de l'insertion contenant le gène *Bru1*

Le cultivar R570, comme tous les cultivars modernes de canne à sucre, est dérivé d'hybridations entre l'espèce domestique productrice de sucre *S. officinarum* et l'espèce sauvage *S. spontaneum*. Des expériences d'hybridation *in situ* ont montré que 80 % des chromosomes de R570 proviennent de *S. officinarum*, 10 % de *S. spontaneum* et 10 % sont issus de recombinaisons ou de translocations entre ces deux espèces (D'Hont *et al.*, 1996). La carte physique partielle de l'haplotype cible de la région de *Bru1* a montré la présence d'une insertion d'une taille indéterminée. La délimitation de l'insertion est basée pour l'instant uniquement sur la rupture de la synténie entre l'haplotype cible et les autres haplotypes hom(é)ologues. Aucune recombinaison n'a été observée à l'intérieur de cette insertion dans la descendance de R570 (Le Cunff *et al.*, 2008). D'après une étude basée sur des marqueurs AFLP (Hoarau *et al.*, 2001), le groupe de coségrégation porteur de l'insertion proviendrait vraisemblablement de *S. officinarum* (Figure 4.1).

Une étude portant sur 404 cultivars de canne à sucre d'origines géographiques différentes, et dont la moitié est résistante à la rouille brune, a montré un fort déséquilibre de liaison dans la région de *Bru1* (Costet *et al.*, en préparation). Aucune recombinaison n'a été retrouvée entre deux marqueurs dérivés de la carte physique de l'haplotype cible (9O20D11/RsaI\* et R12H16\*) présentant une association très forte avec la résistance à la rouille. De plus, ces marqueurs n'ont été observés dans aucun des cultivars sensibles (Costet *et al.*, en préparation). Ces deux marqueurs sont donc considérés comme marqueurs diagnostiques pour la présence de *Bru1* dans les cultivars modernes.

Les objectifs de notre étude sont d'utiliser ces marqueurs pour retracer l'origine de l'insertion dans la généalogie « décrite » du cultivar R570 et déterminer si cette insertion provient de *S. officinarum*, de *S. spontaneum* ou d'une autre espèce du genre *Saccharum*.

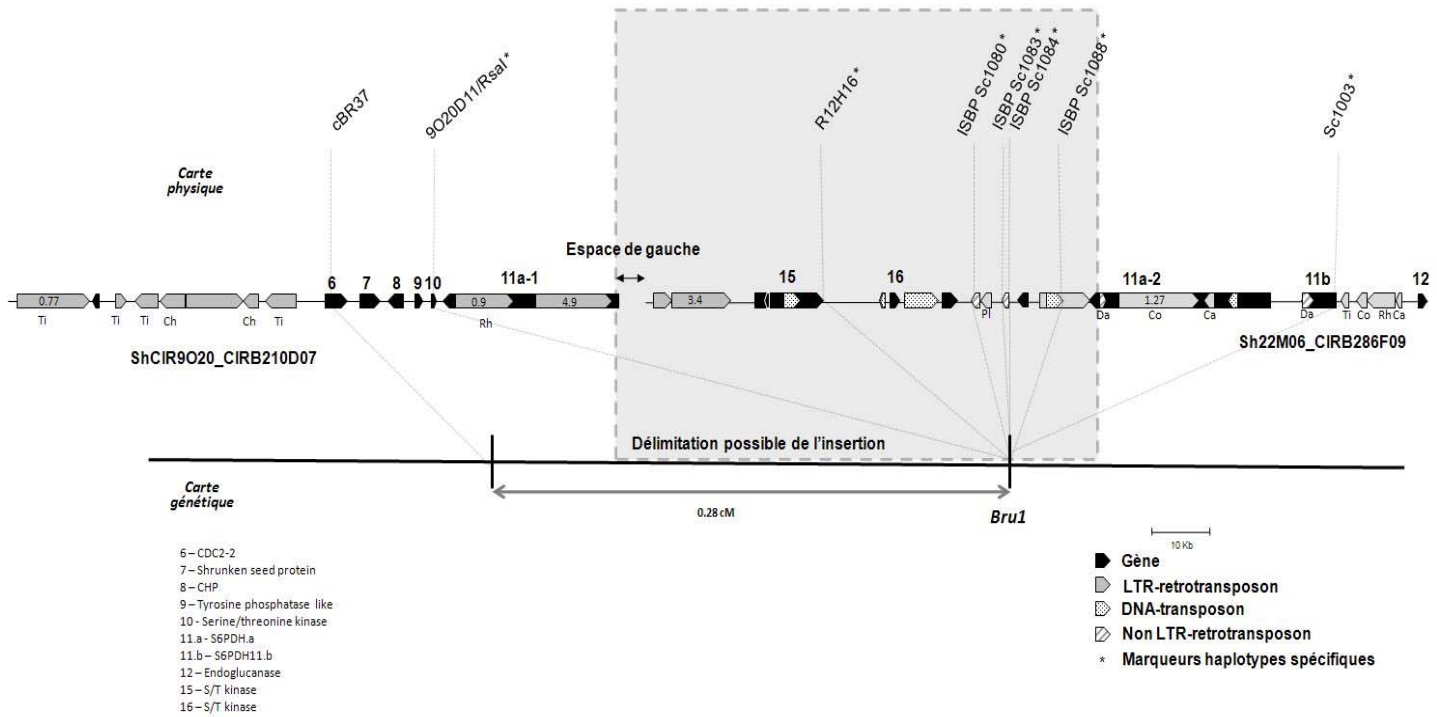


Figure 4.2. Localisation des marqueurs PCR et ISBP sur la séquence de l'haplotype cible de la région de *Bru1*.

## **A. Matériels et Méthodes**

### **A.1. Matériel végétal.**

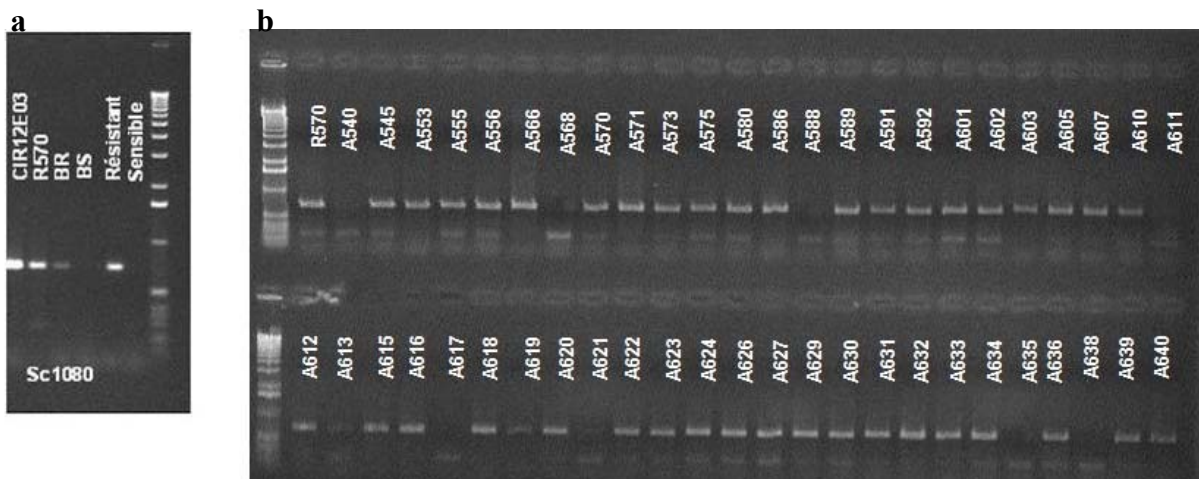
Une collection de 136 clones (avec parfois des clones en plusieurs exemplaires) représentant les différentes espèces du genre *Saccharum* et quelques-uns des premiers hybrides interspécifiques à l'origine des cultivars modernes, dont un certain nombre de clones appartenant à la généalogie du cultivar R570, a été analysée :

- 73 clones de *S. officinarum* (espèce cultivée sucrée)
- 31 clones de *S. spontaneum* (espèce sauvage)
- 13 clones de *S. robustum* (espèce sauvage à partir de laquelle aurait été domestiquée *S. officinarum*)
- 11 clones de *S. barberi* et *S. sinense* (hybrides naturels entre *S. officinarum* et *S. spontaneum*, anciennement cultivés)
- 8 clones représentant quelques-uns des premiers hybrides interspécifiques à l'origine des cultivars modernes

### **A.2. Marqueurs PCR**

#### **A.2.1. Marqueurs PCR dérivés de la carte physique de *Bru1***

La collection des 136 clones de *Saccharum* a été analysée avec quatre marqueurs PCR dérivés de la carte physique de l'haplotype cible (Figure 4.2): le marqueur PCR cBR37\* cartographié à 0,28 cM du gène *Bru1* du côté gauche sur la carte physique, les deux marqueurs diagnostiques (9O20D11/RsaI\* et R12H16\*) qui coségrègent avec *Bru1* et le marqueur Sc1003\* qui coségrège lui aussi avec le gène *Bru1* et situé *a priori* à l'extérieur de l'insertion du côté droit sur la carte physique (chapitre 2, paragraphe C.1).



**Figure 4.3. Exemple de profils PCR obtenus avec le marqueur ISBP Sc1080.**

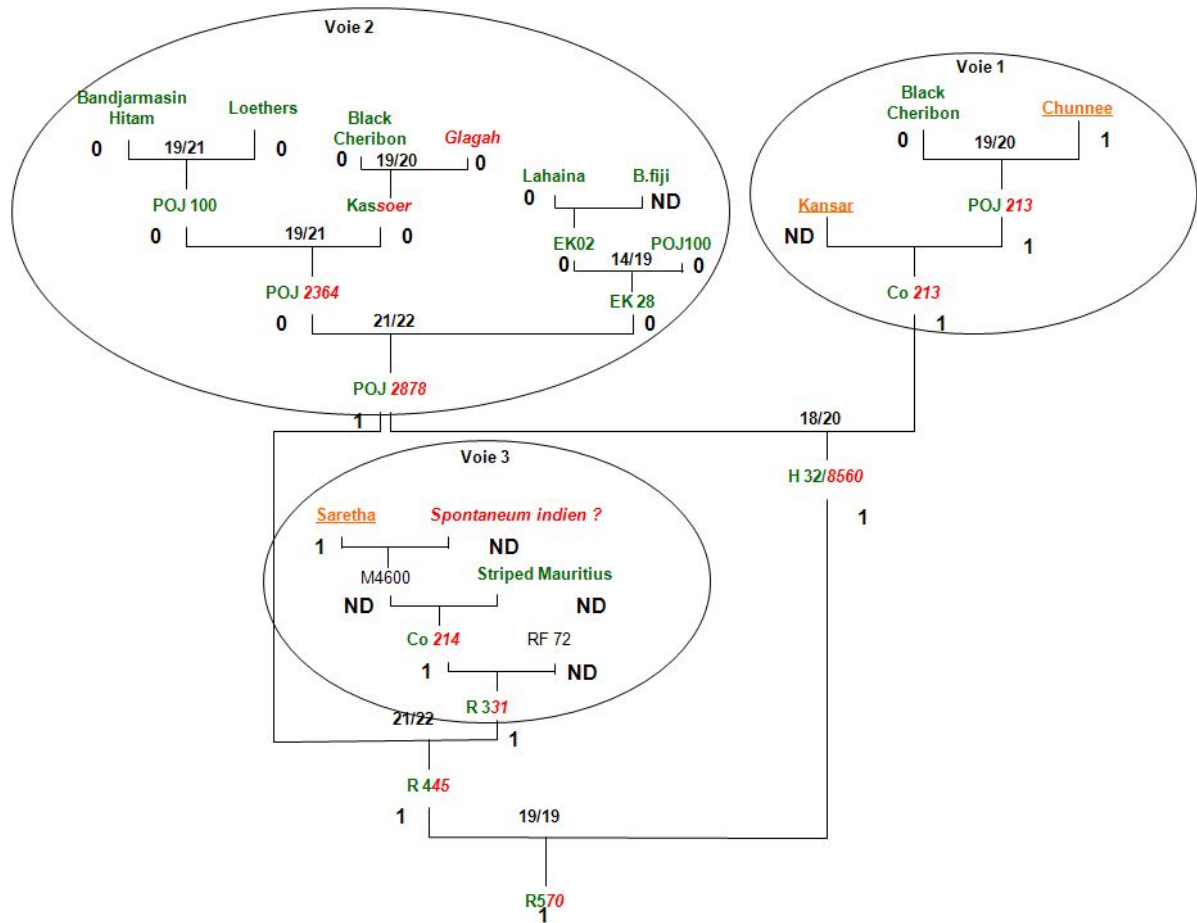
a) Profil PCR du marqueur ISBP Sc1080 lié au caractère sensible/résistant (BR = Bulk Résistant, BS = Bulk Sensible)

b) Profil PCR du marqueur ISBP Sc1080 sur 49 individus de la population de cartographie issue de l'autofécondation de R570

Les réactions PCR sont réalisées dans un volume final de 50 µl avec les composants suivants : 2,5 µL d'ADN génomique à 10 ng.µL<sup>-1</sup> ou 2,5 µL d'ADNc, 5 µL de tampon de réaction 10X (Eurobio), 0,5 µL de MgCl<sub>2</sub> à 25 mM, 4 µL de dNTP à 2 mM, 2,5 µL d'amorces sens et antisens à 10 µM et 1 unité de *Taq* DNA polymerase (Eurobio). Le programme d'amplification PCR commence par une étape de dénaturation de l'ADN à 94°C pendant 4 min puis 35 cycles à 94°C pendant 30s, 55 °C pendant 30s, 72°C pendant 45 s et une extension finale à 72°C pendant 8 min. Les produits PCR sont visualisés sur un gel d'agarose à 2 %.

### A.2.2. Marqueurs ISBP

En nous appuyant sur la séquence et l'annotation des BAC disponibles de l'haplotype cible, nous avons développé de nouveaux marqueurs PCR haplotypes-spécifiques de type ISBP pour « Insertion Site-Based Polymorphism ». Ces marqueurs ciblent les zones d'insertion des éléments transposables dans le génome (Paux *et al.*, 2006). Une amorce est définie dans un élément transposable et l'autre en bordure de cet élément. L'amplicon ainsi obtenu cible potentiellement un site unique d'insertion d'un élément transposable dans le génome de l'espèce étudiée (Paux *et al.*, 2006; Paux *et al.*, 2010). En effet, Garsmeur *et al* (2010) ont montré une absence globale de colinéarité des éléments transposables entre les haplotypes hom(é)ologues de la région de *Bru1* et donc la possibilité d'obtenir une nouvelle source de marqueurs haplotypes-spécifiques. Pour tester ces marqueurs, nous avons recherché un polymorphisme de type résistant/sensible. Les amplifications PCR ont été réalisées sur le BAC où le couple d'amorces a été défini, sur l'ADN génomique de R570, sur l'ADN génomique d'un bulk d'individus résistants à la rouille (BR) et sur l'ADN génomique d'un bulk d'individus sensibles à la rouille (BS). Seuls les marqueurs qui amplifient sur le BAC, R570 et BR, mais pas sur BS, sont conservés. Nous avons défini des marqueurs ISBP pour chaque site d'insertion d'élément transposable présent dans l'insertion de l'haplotype cible entre les marqueurs R12H16\* et Sc1003\*. Dix éléments transposables, dont certains présentent le même site d'insertion, sont présents dans l'insertion et 18 couples d'amorces ont été définis. Nous avons ainsi pu dériver quatre marqueurs ISBP haplotypes-spécifiques : Sc1080\*, Sc1083\*, Sc1084\* et Sc1088\* (Figure 4.3). Ces quatre marqueurs ISBP ont ensuite



**Figure 4.4. Distribution des marqueurs 9O20D11/RsaI\* et R12H16\* dans la généalogie du cultivar R570.**

La présence *versus* absence des marqueurs 9O20D11/RsaI\* et R12H16\* est indiquée par 1 *versus* 0.

Les liens de parenté vérifiés par les analyses microsatellites sont représentés par une fraction : nombre de bandes du descendant présentes chez au moins un des parents / nombre de bandes totales du descendant.

Les clones verts appartiennent à *S. officinarum*, rouge à *S. spontaneum*, orange surligné à *S. barberi* et vert/rouge sont des hybrides entre *S. officinarum* et *S. spontaneum*.



été cartographiés. Ils coségrègent avec le gène *Bru1* (Figure 4.3). Les conditions PCR utilisées sont celles décrites dans Paux *et al.*, (2010). Les produits PCR sont déposés sur un gel d'agarose de 2 ou 4 % suivant la taille des amplifiats.

### **A.2.3. Marqueurs microsatellites**

La collection de clones de *Saccharum* a été génotypée au laboratoire à l'aide de trois microsatellites (CV29, CV37 et CV38) décrits par Maccheroni *et al.* (2007). La combinaison de ces trois marqueurs microsatellites permettait de différencier une collection de 1 205 clones de canne à sucre. Le nombre moyen d'allèles obtenus par génotype avec la combinaison des trois allèles est de 19,6 et le nombre moyen d'allèles discriminatoires est de 13,5. Nous avons utilisé ces données pour comparer le génotype des clones présents en plusieurs exemplaires dans notre collection et tester l'exactitude de la généalogie « décrite » de R570. Les résultats sont présentés dans l'annexe 2.

## **B. Distribution des marqueurs dans la collection**

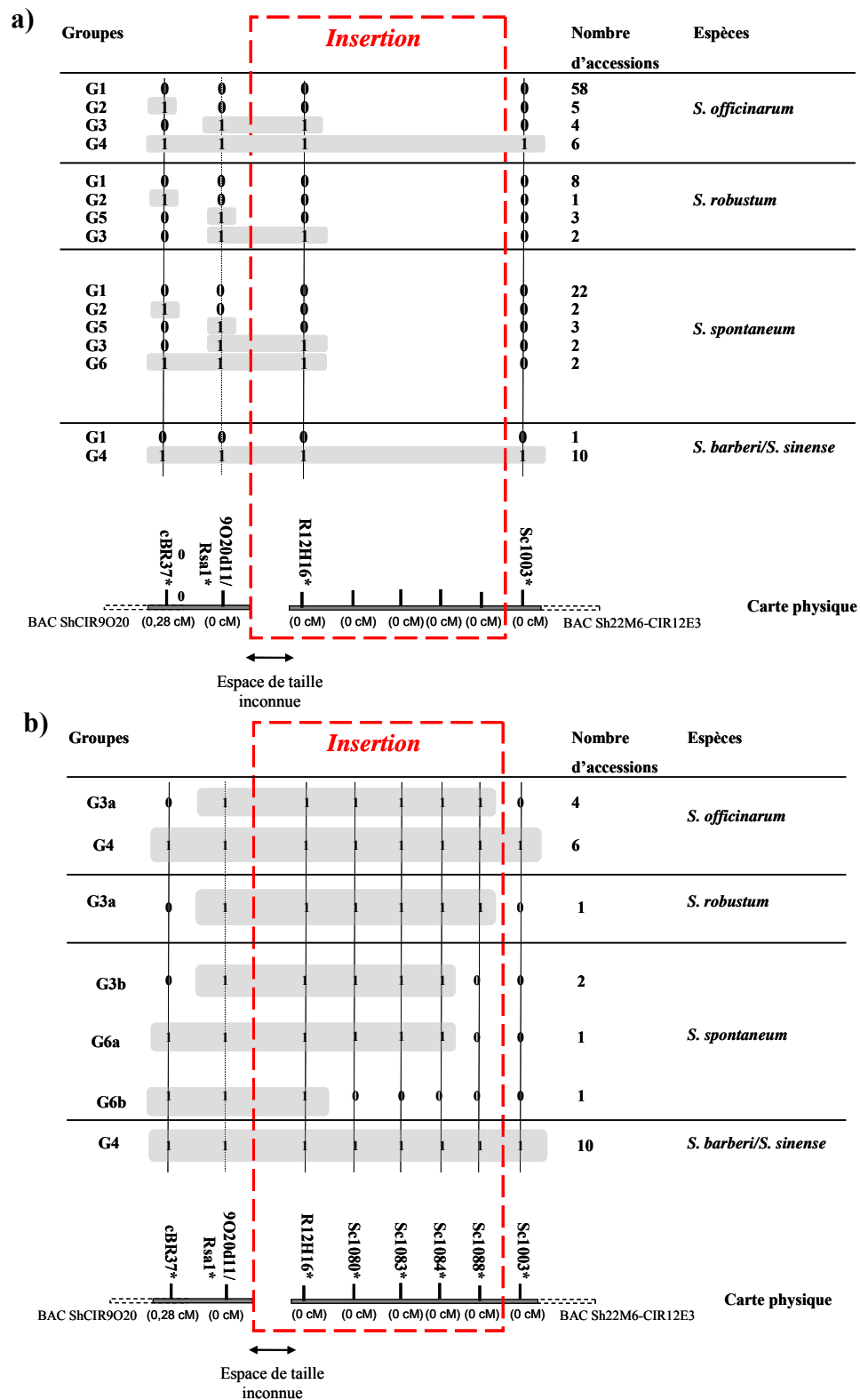
### **B.1. Distribution des marqueurs dans la généalogie du cultivar moderne R570**

Nous avons analysé les marqueurs 9O20D11/RsaI\* et R12H16\* sur tous les clones de la généalogie « décrite » de R570 disponibles. Les résultats sont présentés dans la figure 4.4. Les « parents » (R445 et H 32/8560) et les « grands-parents » (POJ 2878, R331 et Co 213) de R570 possèdent les deux marqueurs diagnostiques. L'insertion a donc pu être transmise à R570 (chez qui elle est en copie unique) par trois voies : l'ascendance de Co 213, l'ascendance de POJ 2878 et l'ascendance de R331. Les résultats seront présentés suivant ces trois voies :



**Voie 1 :** Dans l'ascendance de Co 213, le cultivar POJ 213 et le clone *S. barberi* Chunnee possèdent les deux marqueurs. Le clone *S. officinarum* Black Cheribon ne présente pas les deux marqueurs. Le clone *S. barberi* Kansar n'a pas été analysé. Les profils microsatellites (Annexe 2) sont compatibles avec le fait que les parents de POJ 213 sont Black Cheribon et Chunnee si on tolère la variation d'un allèle au cours du temps : 19 bandes ont été retrouvées chez les parents sur les 20 présentes chez le descendant POJ 213 (Figure 4.4). Dans cette ascendance, l'insertion pourrait donc provenir du clone Chunnee. Chunnee appartient au groupe des *S. barberi* qui est issu d'hybridations naturelles entre les génomes *S. officinarum* et *S. spontaneum* (D'Hont *et al.*, 2002).

**Voie 2 :** Aucun des ascendants présumés de POJ 2878 ne présente les deux marqueurs diagnostiques alors que POJ 2878 les possède (Figure 4.4). Ni les parents de POJ 2878 (POJ 2364-hybrides et EK 28-*S. officinarum*) ni les grands-parents présumés ne possèdent les deux marqueurs diagnostiques. Sur dix clones de POJ 2878 analysés, tous présentent les deux marqueurs diagnostiques mais deux profils microsatellites différents (variation de deux bandes) ont été observés. Nous avons analysé aussi six clones de POJ 2364 dont aucun ne présente les deux marqueurs diagnostiques. Nous avons observé deux profils microsatellites différents (variation d'une bande). Sur les deux clones d'EK 28 analysés, aucun ne possède les marqueurs diagnostiques. Les analyses microsatellites semblent confirmer que les parents décrits pour POJ 2878 sont bien EK 28 et POJ 2364. En effet, 21 bandes ont été retrouvées chez les parents sur les 22 présentes chez le descendant POJ 2878 (Figure 4.4). Toutefois, le fait que POJ 2878 possède les marqueurs diagnostiques alors que ses parents ne les possèdent pas n'est pas compatible avec cette généalogie. De même, les parents et grands-parents de POJ 2364 semblent compatibles avec les données microsatellites. Par contre, EK 28 ne proviendrait pas du croisement entre EK 02 et POJ 100 : 14 bandes ont été retrouvées chez les parents sur les 19 présentes chez le descendant EK 28. Il est intéressant de noter que la bande supplémentaire de POJ 2878 qui ne se retrouve chez aucun des ascendants de POJ 2878 se retrouve par contre chez tous les clones de la généalogie de R570 présentant les deux marqueurs diagnostiques. Cette bande a été analysée sur la descendance issue de l'autofécondation de R570 afin de déterminer si elle pourrait être liée au caractère de résistance à la rouille mais celle-ci ne ségrège pas dans la descendance et ne peut pas être cartographiée. Cette bande a été recherchée aussi dans deux populations utilisées dans des



**Figure 4.5. Profil des marqueurs PCR et ISBP sur la collection de clones de canne à sucre.**

**a)** Profils des 4 marqueurs PCR sur la collection.

**b)** Profils des 4 marqueurs PCR et des X marqueurs ISBP sur la collection.

Le chiffre 1 correspond à la présence du marqueur et le chiffre 0 correspond à l'absence du marqueur.

études de déséquilibre de liaison phénotypées pour la rouille. Mais aucune corrélation n'est apparue entre la résistance à la rouille et la présence de cette bande. Le parent POJ 2878 possède donc l'insertion mais sa généalogie est incertaine.

**Voie 3** : Nous ne disposons pas de la plupart des clones présents dans l'ascendance de R331. Toutefois, nous avons montré que deux des ascendants de R331, Saretha et Co 214 possèdent les deux marqueurs diagnostiques de l'insertion. Le clone Saretha appartient comme Chunnee à l'espèce *S. barberi*.

## **B.2. Distribution des marqueurs dans la collection de clones des espèces ancestrales.**

Les résultats obtenus sur la collection de 136 clones de *Saccharum* avec les quatre marqueurs PCR liés à *Bru1* chez R570 sont présentés dans la figure 4.5.

Pour les clones *S. officinarum*, quatre groupes ont pu être distingués sur la base de leur profil PCR. Deux groupes présentent au moins une partie de l'insertion (G3 et G4). Six clones sur 73 possèdent les quatre marqueurs (G4) et quatre clones ne possèdent que les marqueurs 9O20D11/RsaI\* et R12H16\* (G3), soit un total de 10 clones sur 73 ayant au moins une partie de l'insertion (13 % des *S. officinarum*).

Pour les clones *S. robustum*, quatre groupes ont été distingués dont un seul montrant la présence d'une partie de l'insertion (G3). Seul un clone sur 13 a présenté l'insertion, soit 7 % des clones *S. robustum*. Aucun des clones de *S. robustum* analysés ne comporte le marqueur Sc1003\*.

Pour les clones *S. spontaneum*, cinq groupes ont pu être identifiés dont deux groupes montrant la présence d'au moins une partie de l'insertion : G3 et G6 (présence des marqueurs cBR37, 9O20D11/RsaI\* et R12H16\*). Ces deux groupes regroupent 4 clones sur 31 soit 13 % des *S. spontaneum*. Nous avons observé, comme pour *S. robustum*, qu'aucun des clones de *S. spontaneum* analysés ne possédait le marqueur Sc1003\*.

**Tableau 4.1. Profils PCR et nombre de profils microsatellites de 26 clones *Saccharum***

Espèce	Sous- groupe	Origine géographique	Nom	Nombre de chromo- somes*	Nombre de profils SSR : Nbre bandes variantes /	Marqueurs PCR **								Nombre de tubes / origine des tubes si profils
						cBR37.7 (0.28 cM)	9O20D11 / RsaI (0 cM)	R12H16 (0 cM)	Sc1080 (0 cM)	Sc1083 (0 cM)	Sc1084 (0 cM)	Sc1088 (0 cM)	Sc1003 (0 cM)	
<i>S. officinarum</i>		Nouvelle Guinée	NG 28 288	80	1 profil	+	+	+	+	+	+	+	+	1 tube
<i>S. officinarum</i>		Fiji	Fiji 24	80	1 profil	0	+	+	+	+	+	+	0	5 tubes
<i>S. officinarum</i>		Fiji	Fiji 15	80	1 profil	+	+	+	+	+	+	+	+	2 tubes
<i>S. officinarum</i>		Indonésie / île Maluku	IM 76 244	-	1 profil	+	+	+	+	+	+	+	+	1 tube
<i>S. officinarum</i>		Indonésie / île Sulawesi	IS 76 116	-	1 profil	+	+	+	+	+	+	+	+	2 tubes
<i>S. officinarum</i>		Nouvelle Calédonie îles	NC 24	95	1 profil	0	+	+	+	+	+	+	0	2 tubes
<i>S. officinarum</i>		Mauriciennes Nouvelle Guinée	Bambou rose	80	1 profil	+	+	+	+	+	+	+	+	1 tube
<i>S. officinarum</i>		Indonésie / île Irian Jaya	NG 57 198	80	1 profil	0	+	+	+	+	+	+	0	1 tube
<i>S. officinarum</i>		Nouvelle Guinée	IJ 76 291	-	1 profil	0	+	+	+	+	+	+	0	1 tube
<i>S. officinarum</i>		Nouvelle Guinée	NG 77 66	-	1 profil	+	+	+	+	+	+	+	+	2 tubes
<i>S. robustum</i>		Nouvelle Guinée	NG 28 251	80	1 profil	0	+	+	+	+	+	+	0	2 tubes
<i>S. spontaneum</i>		Hawaï / île Molokai	Mol 5801	80	1 profil	0	+	+	+	+	+	0	0	3 tubes
<i>S. spontaneum</i>		Thaïlande	US 56 00193	88	2 profils : 2/15	0 0	+	+	+	+	+	0 0	0 0	2 tubes / origine inconnue
<i>S. spontaneum</i>		Kenya	Isiolo	124	1 profil	+	+	+	+	+	+	0	0	3 tubes
<i>S. spontaneum</i>		Inde	Co Local	64	1 profil	+	+	+	0	0	0	0	0	4 tubes
<i>S. barberi</i>	Saretha katha	Inde	Chunnee	91	2 profils : 4/27	+	+	+	+	+	+	+	+	14 tubes / Afrique du Sud, Guadeloupe et la Réunion
<i>S. barberi</i>	Saretha katha	Inde	Saretha	91-92	1 profil	+	+	+	+	+	+	+	+	8 tubes
<i>S. barberi</i>	Sunnabile Mojorah	Inde	Hullu kabbu Rakhra /	106	1 profil	+	+	+	+	+	+	+	+	5 tubes
<i>S. barberi</i>	Sunnabile Dhaulu	Inde	Rathara	82	1 profil	+	+	+	+	+	+	+	+	5 tubes
<i>S. barberi</i>	Sunnabile Dhaulu	Inde	Pathri	82	1 profil	+	+	+	+	+	+	+	+	4 tubes
<i>S. barberi</i>	Sunnabile Dhaulu	Inde	Paunra	82	1 profil	+	+	+	+	+	+	+	+	4 tubes
<i>S. barberi</i>	Mungo	Inde	P. Mango***	82	1 profil	0	0	0	0	0	0	0	0	2 tubes
<i>S. sinense</i>	Pansahi	Japon	Oshima	117	1 profil	+	+	+	+	+	+	+	+	8 tubes
<i>S. sinense</i>	Pansahi	Inde / région Bihar	Puri	116	1 profil	+	+	+	+	+	+	+	+	4 tubes
<i>S. sinense</i>	Pansahi	Chine	UBA	116-118	1 profil	+	+	+	+	+	+	+	+	8 tubes
<i>S. sinense</i>	Pansahi	Chine	UBA naguin	117	1 profil	+	+	+	+	+	+	+	+	7 tubes

\* d'après Price, (1957); Price, (1968) ; Jagathesan *et al*, (1970) ; D'hont *et al*, (1993) ; Jannoo *et al*, (1999) ; D'Hont *et al*, (2002).

\*\* marqueurs PCR spécifiques de l'haplotype cible porteur de l'insertion chez le cultivar moderne R570. La distance exprimée en cM correspond à la distance génétique par rapport au gène *Brul* chez R570

\*\*\* seul clone *S. barberi* analysé de la collection ne présentant pas l'insertion

La zone en gris correspond aux marqueurs présents dans l'insertion

Pour les clones *S. barberi* / *S. sinense*, nous avons observé deux groupes : un groupe avec les quatre marqueurs PCR (G4) et un groupe avec aucun des quatre marqueurs (G1). Dix clones sur 11 présentent les quatre marqueurs (G4), soit 90 % des clones *S. barberi* / *S. sinense*.

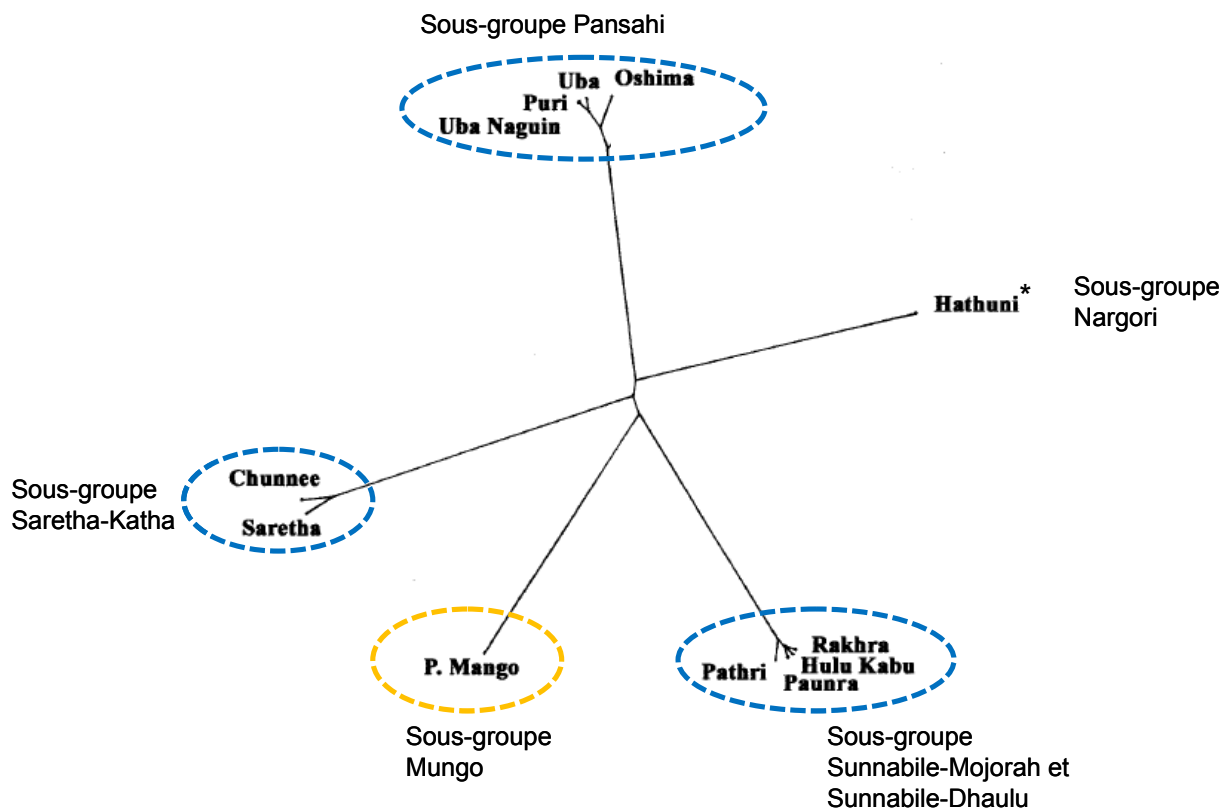
Pour affiner l'analyse sur les clones de la collection de *Saccharum*, quatre marqueurs ISBP situés entre le marqueur R12H16\* et Sc1003\* (Figure 4.2) ont été utilisés sur les clones montrés comme possédant au moins une partie de l'insertion : 10 clones pour *S. officinarum*, 1 clone pour *S. robustum*, 4 clones pour *S. spontaneum* et 10 clones pour *S. barberi* / *S. sinense*. Les résultats obtenus sont présentés dans la figure 4.5. Tous les clones de *S. officinarum*, *S. robustum* et *S. barberi* / *S. sinense* possèdent les quatre marqueurs ISBP. Pour *S. spontaneum*, nous avons identifié trois groupes (Figure 4.5):

- Groupe 3b : 2 clones avec les marqueurs 9O20D11/RsaI\*, R12H16\*, Sc1080\*, Sc1083\* et Sc1084\* mais pas le marqueur Sc1088\*.
- Groupe 6a : 1 clone avec les marqueurs cBR37, 9O20D11/RsaI\*, R12H16\*, Sc1080\*, Sc1083\* et Sc1084\* mais pas le marqueur Sc1088\*.
- Groupe 6 : 1 clone avec les marqueurs cBR37, 9O20D11/RsaI\* et R12H16\* mais aucun des marqueurs ISBP.

Les groupes observés nous ont indiqué que les clones *S. officinarum*, *S. robustum* et *S. barberi* / *sinense* semblent contenir l'insertion dans sa totalité ou pas du tout. Pour *S. spontaneum*, les résultats nous indiquent que ceux-ci ne contiennent qu'une partie de l'insertion dont la taille varie suivant les clones *S. spontaneum* étudiés.

### **C. Conclusion sur l'origine de l'insertion**

L'analyse de la généalogie « décrite » du cultivar R570, bien que seulement partiellement compatible avec les résultats du marquage moléculaire, tend à montrer que l'insertion pourrait avoir été transmise par un clone appartenant à l'espèce *S. barberi* :



**Figure 4.6. Phylogénie de 13 clones de *S. barberi* et *S. sinense* basée sur 203 marqueurs RFLP révélés par une combinaison de 10 sondes / enzyme (d'après D'Hont *et al.*, 2002).**

Le rond en pointillé bleu indique un sous-groupe portant l'insertion. Le rond en pointillé jaune indique un sous-groupe ne possédant pas l'insertion.

\* clone mal étiqueté dans l'étude de D'Hont *et al.*, 2002



Chunnee ou Saretha. Ces deux *S. barberi* semblent posséder l'ensemble de l'insertion (présence de tous les marqueurs). De plus, un troisième clone *S. barberi*, Kansar, qui n'a pu être analysé, intervient dans la généalogie de R570. Le clone Kansar appartient au même sous-groupe de *S. barberi* que les clones Saretha et Chunnee qui possèdent l'insertion (sous-groupe Saretha-katha) (Bremer, 1966). L'analyse de la collection de *Saccharum* a montré que l'insertion était présente en faible proportion (13-14%) chez les espèces *S. officinarum* (10 clones sur 73), *S. spontaneum* (4 clones sur 36) et en très faible proportion chez *S. robustum* (1 clone sur 13). Par contre, elle a été retrouvée en proportion beaucoup plus importante (90 %) chez le groupe *S. barberi* / *S. sinense* (10 clones sur 11) (Tableau 4.1).

Les clones « barberi » et « sinense » ont été classés en deux espèces par certains botanistes (Bremer, 1932; Price, 1968) et divisés en 7 sous-groupes sur la base de caractères morphologiques et leur nombre de chromosomes (Paton *et al.*, 1978; Daniels *et al.*, 1991). Toutefois, il a été montré par marquages moléculaires (RFLP et GISH) qu'il ne s'agissait pas à proprement parler d'espèces mais de groupes de clones issus d'hybridations interspécifiques entre les génomes *S. officinarum* et *S. spontaneum* (D'Hont *et al.*, 2002). D'autre part, l'analyse au moyen des marqueurs moléculaires a montré que ces sous-groupes correspondent à des variants somaclonaux (D'Hont *et al.*, 2002). Nous avons analysé 11 représentants appartenant à 5 de ces sous-groupes. L'insertion a été retrouvée chez tous les clones analysés des sous-groupes Saretha-Katha, Sunnabile-Dhaultu, Sunnabile-Mojorah et Pansahi mais pas chez l'unique représentant du sous-groupe Mungo (Figure 4.6).

Chez les espèces *S. officinarum*, *S. robustum* et *S. barberi* / *S. sinense*, tous les marqueurs de l'insertion sont retrouvés et donc ces clones semblent contenir l'ensemble de l'insertion alors que chez *S. spontaneum*, les clones présentent des parties variables de l'insertion. Une interprétation pourrait être que cette région correspondant à l'insertion peut recombiner dans un contexte *S. spontaneum* mais pas dans le contexte des autres espèces.



---

## **Chapitre 5 : Discussion générale et perspectives**

---



## Chapitre 5 : Discussion générale et perspectives

Cette thèse s'est articulée autour de trois objectifs principaux : (I) compléter la carte physique de l'haplotype cible contenant le gène *Bru1* de résistance à la rouille chez le cultivar moderne R570 de canne à sucre, (II) identifier les gènes candidats présents sur la partie séquencée de l'haplotype cible et (III) rechercher l'origine du fragment de chromosome inséré dans cet haplotype porteur du gène *Bru1*. Les deux premiers objectifs nous ont permis d'avancer vers le clonage de *Bru1*. La couverture d'un des espaces et l'annotation fine de celui-ci indiquent que *Bru1* se situerait dans l'insertion présente sur l'haplotype cible. Le troisième objectif nous a quant à lui permis d'avancer dans la compréhension de l'origine de cette insertion.

### **A. Cartographie physique de l'haplotype porteur du gène *Bru1***

Les contraintes majeures d'un projet de clonage positionnel chez la canne à sucre sont son haut niveau de polyploïdie qui induit une taille de génome (10 Gb) et une redondance ( $2n = 12 X$ ) très importantes.

La forte redondance implique d'utiliser des stratégies très lourdes pour la réalisation de cartes génétique et physique d'un haplotype spécifique parmi plus de dix autres. Pour contourner cette contrainte, une stratégie exploitant des espèces diploïdes apparentées a été utilisée avec succès chez d'autres espèces polyploïdes en particulier chez le blé où plusieurs gènes de résistance ont été clonés grâce à cette stratégie (Feuillet *et al.*, 2003; 2003; Yahiaoui *et al.*, 2004). Cette stratégie ne peut être appliquée à la canne à sucre parce qu'il n'existe aucune espèce diploïde dans le genre *Saccharum*.

Pour avancer dans le clonage de *Bru1* chez la canne à sucre, Le Cunff *et al* (2008) ont utilisé les relations synténiques existantes entre la canne à sucre et deux espèces diploïdes de genres proches, le sorgho et le riz. Cette stratégie a été fructueuse et a permis de développer une carte génétique fine de la région entourant *Bru1* et une carte physique partielle de l'haplotype cible. Toutefois, la présence d'une insertion spécifique à l'haplotype cible a empêché de poursuivre cette stratégie pour finaliser la carte physique.



dans une moindre mesure, avec les régions orthologues du sorgho et du riz (Garsmeur *et al.*, en presse). Dans notre étude, nous avons exploité cette très bonne conservation des gènes entre les différents haplotypes hom(é)ologues pour combler l'espace de droite de la carte physique de l'haplotype cible. Le fait que l'haplotype cible présente une région absente des autres haplotypes hom(é)ologues chez le cultivar R570 n'a pas permis d'exploiter cette stratégie pour l'espace de gauche. Aucune région orthologue à l'insertion chez le sorgho et le riz n'a pu être identifiée. De la même façon, des gènes de résistance à la rouille chez le blé *Lr10* (Feuillet *et al.*, 2003), *Lr21* (Huang *et al.*, 2003) et *Pm3* (Yahiaoui *et al.*, 2004) et le gène *Rpg1* chez l'orge (Brueggeman *et al.*, 2002) ne présentent pas d'orthologue dans le génome du riz. Seules les techniques classiques de marche sur chromosome peuvent donc être utilisées pour combler l'espace de gauche.

La grande taille du génome est une autre contrainte qui implique lors des étapes de cartographie physique de manipuler un grand nombre de clones BAC. Pour construire une banque BAC correspondant à une couverture du génome totale de 5 X avec une moyenne d'inserts de 100 Kb, le nombre de clones BAC pour un cultivar moderne de canne à sucre comme R570 est de 500 000. En comparaison, chez le sorgho, 80 000 clones BAC suffisent.

Pour construire la première carte physique de la région de *Bru1*, Le Cunff *et al.*, (2008) ont utilisé deux banques BAC représentant 2,7 fois le génome total de R570 et 4,1 fois la région de *Bru1*. Cette couverture étant un peu faible, nous avons entrepris d'augmenter le taux de couverture en produisant de nouveaux clones BAC avec une nouvelle enzyme de restriction pour éviter les biais de représentation dû à l'enzyme de clonage. Pour enrichir la banque en haplotype cible, nous avons utilisé, comme Le Cunff *et al.*, (2008), l'ADN de descendants issus de l'autofécondation du cultivar R570 présentant deux copies de la région cible. Cette banque comprend 119 040 clones BAC ce qui correspond à une couverture représentant 1,6 fois le génome total et 3,2 fois la région de *Bru1*. En cumulant les trois banques BAC, la région cible est représentée statistiquement 7,3 fois. Cette couverture n'est pas encore très importante mais des couvertures légèrement inférieures ont permis d'identifier des gènes recherchés. Par exemple, le clonage chez l'orge des gènes *Rpg1*, *Rpg5* et du gène *Rph7* a été réalisé avec une banque d'une couverture du génome totale de 6,3 X (Brueggeman *et al.*, 2002; Brunner *et al.*, 2003; Brueggeman *et al.*, 2008). Le gène *Vrn1* chez le maïs a été cloné avec une banque d'une couverture du génome total de 5,6 X (Yan *et al.*, 2003).

Cette approche a été fructueuse car elle nous a permis de combler un des deux espaces de la carte physique de l'haplotype cible et d'avancer dans le deuxième. Toutefois, la





couverture reste encore modeste, la production de clones BAC supplémentaires nous semble donc une bonne stratégie pour poursuivre le travail visant à combler l'espace non couvert de l'haplotype cible.

Il a été montré que certaines séquences d'organisme eucaryote sont inclonables dans un système procaryote comme *Escherichia coli* (Kang & Cox, 1996). La nature de telles séquences n'est pas bien connue mais certaines ont déjà été décrites : la présence de longue région répétée, des séquences riches en A-T ou des séquences capables de former des structures secondaires anormales comme l'ADNz (ADN s'enroulant dans le sens inverse) (Kang & Cox, 1996; Ravin & Ravin, 1999; Razin *et al.*, 2001). La réalisation de banque BAC passe par la transformation de cellule d'*E. coli*. Donc, même avec une couverture du génome totale suffisante, un tel type de région ne sera jamais représenté dans une banque BAC. Bien que nous n'ayons pas d'éléments pour penser que nous soyons dans une telle région, cette hypothèse ne peut être exclue. L'utilisation des levures (organisme eucaryote) permettrait de contourner ce problème car elles procureraient une meilleure stabilité de ce type de séquence dans les vecteurs de transformation (Gardner *et al.*, 2002). Ceci impliquerait d'utiliser des banques YAC (Yeast Artificial Chromosome) avec des inserts beaucoup plus grands que ceux des banques BAC.

## **B. Gènes candidats**

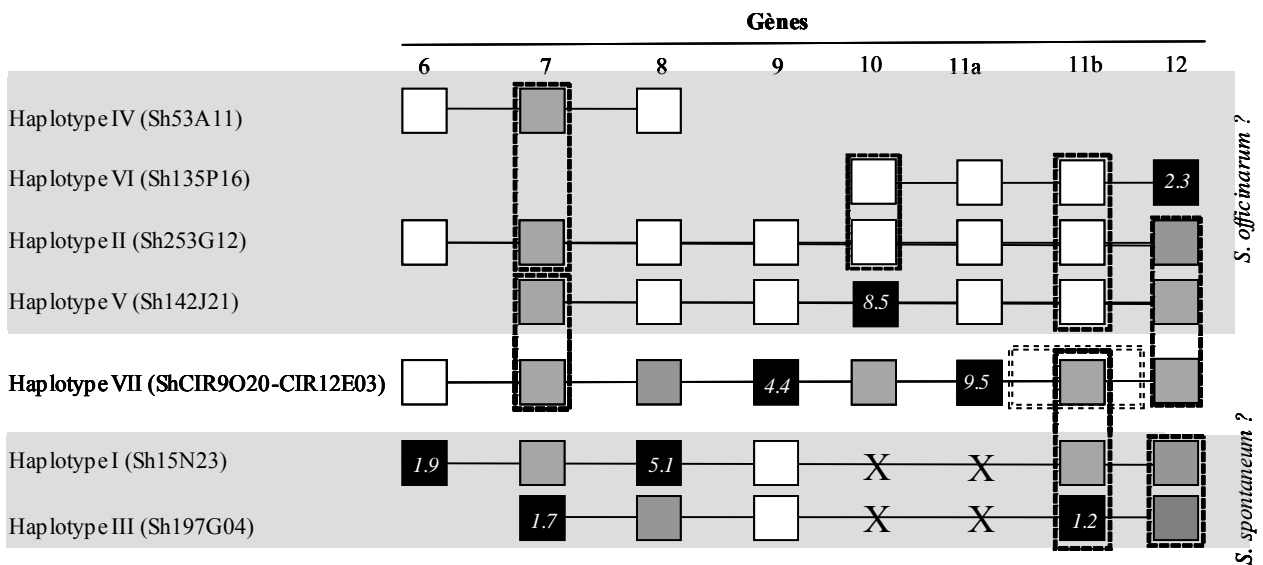
Le gène *Bru1* confère une résistance durable à la rouille brune causée par le champignon pathogène *Puccinia melanocephala*. Asnaghi *et al.*, (2001) ont montré que le gène *Bru1* confère une résistance sur un ensemble d'isolats testés provenant de huit zones géographiques différentes (Brésil, Colombie, Floride, Guadeloupe, La Réunion, Zimbabwe). De plus, le cultivar R570 est résistant à la rouille depuis plus de trente ans dans diverses zones géographiques où il est cultivé : La Réunion, Maurice, Guadeloupe, en Afrique de l'Ouest (Burkina Faso, Gabon et Congo) et de l'Est (Tanzanie, Malawi, Kenya, Swaziland, Mozambique). Une étude portant sur la distribution du gène *Bru1* dans une population de 404 cultivars sélectionnés dans différentes régions du monde indique que *Bru1* est présent chez 87 % des 200 individus résistants (Costet *et al.*, en préparation). La forte représentation de cet haplotype chez les cultivars résistants semble indiquer que ce gène constitue une source majeur de résistance à la rouille brune chez les cultivars modernes et ce depuis longtemps.



L'annotation fine de la région cible nous a permis d'identifier trois S/T kinases présentant une forte homologie avec la protéine codée par le gène *Rpg1* chez l'orge qui intervient dans la résistance à *Puccinia graminis*. Une seule de ces kinases (codée par le gène 15) semble fonctionnelle (présence des douze sous-domaines caractérisant les kinases), avec une structure composée de 6 exons et un très grand intron de plus de 11 kb pour une taille génomique de 12,8 kb. Brueggeman *et al*, (2006) ont identifié plusieurs gènes codant pour des protéines de la famille de *Rpg1*, dont plusieurs qui présentent un grand intron. Chez l'orge, un de ces gènes présente un très grand intron entre l'exon 3 et l'exon 4, comme le gène candidat chez R570. Toutefois, nous n'avons pour l'instant pas pu mettre en évidence l'expression de ce gène dans des cellules non infectées. Ces similitudes avec le gène *Rpg1* sont intrigantes. D'autre part, tout comme *Bru1*, le gène *Rpg1* confère une résistance durable dans le temps qui n'a pas été contournée depuis près de 70 ans (Staples, 2003). Le gène *Rpg1* confère une résistance à plusieurs races d'un même pathogène. Concernant *Puccinia melanocephala*, le pathogène responsable de la rouille brune de la canne à sucre, l'existence de différentes races a été suggéré mais pas clairement démontré (Dean & Purdy, 1984). La source de résistance *Rpg1* proviendrait d'une sélection réalisée sur une variété unique Wisconsin37 et d'un cultivar importé de Suisse en Amérique du Nord (Steffenson, 1992). Les similitudes décrites ci-dessus entre les gènes *Rpg1* et *Bru1* font de la S/T kinases (codée par le gène 15) un candidat intéressant pour *Bru1*.

### **C. Origine de l'insertion**

Au cours de la construction de la carte physique de la région contenant *Bru1*, Le Cunff *et al* (2008) ont identifié des clones BAC correspondant à sept haplotypes hom(é)ologues dont un correspond à l'haplotype cible. Cet haplotype comprend une région supplémentaire par rapport aux autres haplotypes. Les auteurs ont attribué cette région à l'insertion d'un fragment de chromosome de taille indéterminée. Hoarau *et al*, (2001) ont suggéré que le chromosome portant *Bru1* proviendrait du génome de *S. officinarum* : des marqueurs AFLP cartographiés sur le groupe de coségrégation porteur de l'insertion ont été attribués à *S. officinarum*. Certaines sondes RFLP dérivées de l'insertion ont révélé un profil particulier avec une seule bande contrairement aux multiples bandes généralement observées chez ce haut polyploïde. Les analyses de diversité réalisées au moyen de marqueurs moléculaires ont permis de



**Figure 5.1 Représentation schématique de la différenciation des haplotypes de canne à sucre basée sur la comparaison des séquences des allèles hom(é)ologues d'après Garsmeur *et al*, 2010.**

Chaque allèle est représenté par un carré. Pour chaque locus, l'allèle le plus divergent est marqué en noir et son temps de divergence théorique (estimation la plus élevée observée en Million d'année) est indiqué en italique.

Tous les allèles dont les valeurs sont inférieures à un tiers du temps de divergence maximale sont marqués par des carrés blancs. Lorsque les arbres phylogénétiques ont été réalisés, les allèles d'une même branche ont été placés dans une boîte en pointillé verticale, ce qui donne des indices pour la classification (loci 7, 11b et 12)..

Le «x» indique l'absence du gène.

La boîte en pointillée horizontale indique la présence d'une insertion.

clairement différencier les chromosomes des espèces *S. officinarum* et *S. spontaneum* (Lu *et al.*, 1994a) et ont montré que cette dernière espèce présente une diversité moléculaire très importante. Les chromosomes de *S. spontaneum* sont présents en faible proportion chez les cultivars modernes (10 à 20% du génome, D'Hont *et al.*, 1996). A partir de l'ensemble de ces éléments, Le Cunff *et al.* (2008) ont suggéré que cette insertion pourrait provenir de l'espèce *S. spontaneum*, espèce utilisée il y a plus d'un siècle dans les programmes d'amélioration de la canne à sucre pour ses résistances aux maladies. Une autre source de résistance à la rouille conférée par le gène *Bru2* a été caractérisée dans le clone MQ76-53. Ce clone est issu du croisement entre un cultivar moderne et un *S. spontaneum* (Raboin *et al.*, 2006). *Bru2* a été localisé sur un groupe d'homologie (VIII) différent de celui du gène *Bru1* (VII) et serait associé à un groupe de coségrégation provenant de *S. spontaneum* (Raboin *et al.*, 2006).

Le séquençage et l'analyse comparée des BACs correspondant aux sept haplotypes hom(é)ologues ont permis de proposer une délimitation de l'insertion sur la base de la rupture de synténie entre l'haplotype cible et les autres haplotypes hom(é)ologues (Garsmeur *et al.*, 2010). D'autre part, la comparaison du contenu en gènes et la comparaison des séquences des allèles des différents haplotypes ont permis de proposer une hypothèse concernant l'origine spécifique des différents haplotypes hom(é)ologues (Garsmeur *et al.*, 2010 ; cette thèse). Les gènes 10 et 11a sont présents sur tous les haplotypes sauf le I et le III. Globalement, les allèles les plus divergents pour les autres gènes sont retrouvés sur les haplotypes I et III et une partie de l'haplotype VII et les allèles les plus proches se retrouvent sur les haplotypes II, IV, V et VI (Figure 5.1). L'interprétation la plus simple est qu'il existe deux groupes d'haplotypes qui pourraient provenir de deux génomes distincts : *S. officinarum* et *S. spontaneum*. Sur la base de l'existence d'une plus grande diversité au sein de *S. spontaneum* que chez *S. officinarum* (Lu *et al.*, 1994a), il a été proposé que les quatre haplotypes les moins divergents auraient une origine *S. officinarum* (haplotype II, IV, V, VI), les deux haplotypes les plus divergents auraient une origine *S. spontaneum* (haplotype I et III). L'haplotype VII (haplotype cible) présente une structure particulière qui suggère une origine plus complexe. En effet, les allèles des gènes 6, 7 et 12 sont plus proches de ceux des haplotypes attribués à l'espèce *S. officinarum* tandis que l'allèle du gène 11b est plus proche de ceux des haplotypes attribués à l'espèce *S. spontaneum*. Les allèles 9 à 11a de l'haplotype VII sont eux très divergents par rapport à tous les autres.



Costet *et al*, (en préparation) ont montré que l'insertion est fortement présente au sein des cultivars modernes (87% des 200 cultivars résistants analysés dans une collection de 400 cultivars représentative de la diversité mondiale). Il est à noter qu'aucune recombinaison au niveau de cette insertion n'a été identifiée dans cette collection de 400 cultivars et dans une descendance de plus de 2000 individus issus de l'autofécondation de R570 (Costet *et al*, en préparation ; données non publiées). Il semble donc que cette insertion/translocation corresponde à un événement unique qui *a priori* ne recombine pas au sein des cultivars modernes. Ces données suggèrent aussi que cette insertion n'est pas récente et a pu se produire ou être transmise au niveau d'un des clones de la généalogie des cultivars modernes qui a été fortement utilisé en sélection à travers le monde.

Notre travail pour suivre l'origine de l'insertion dans la généalogie décrite de R570 est incomplet et imparfait puisque i) nous n'avons pu analyser tous les individus mentionnés dans cette généalogie, ii) nous savons qu'il existe des erreurs dans les généalogies et nous avons d'ailleurs mis en évidence une incohérence dans une des voies de cette généalogie, iii) nous savons aussi qu'il existe de nombreuses erreurs d'« étiquettes » dans les collections. L'utilisation d'une combinaison de trois marqueurs microsatellites nous a permis d'obtenir des « passeports moléculaires » pour chaque clone et ainsi de limiter ces problèmes en permettant de tester les liens de parenté et d'identifier un certain nombre d'erreurs d'échantillonnage.

Ce travail nous a permis de suggérer que cette insertion pourrait avoir été transmise aux cultivars modernes *via* des clones de l'espèce *S. barberi*. La généalogie des cultivars modernes repose sur un nombre restreint de clones. Seuls 19 clones de *S. officinarum*, 3 clones de *S. spontaneum*, 4 clones de *S. barberi* et un clone de *S. robustum* auraient été utilisés dans les programmes d'introggression du début du siècle dernier qui sont à l'origine de tous les cultivars modernes (Arceneaux, 1965). Parmi eux, sept clones : Black Cheribon, Bandjarmasin Hitam, Loethers, Cristalina (4 *S. officinarum*), Chunnee (1 *S. barberi*), et 2 clones appartenant à *S. spontaneum* provenant d'Inde et de Java sont impliqués de façon bien plus importante que les autres (Arceneaux, 1965). Six de ces clones sont impliqués dans la généalogie décrite de R570 et parmi les 6 que nous avons analysé seul le clone Chunnee possède l'insertion. Toutefois l'incohérence que nous avons trouvé dans la généalogie de R570 concernant le clone POJ 2878, qui possède l'insertion mais pas ses ascendants,





ne nous permet pas de conclure complètement quant au(x) clone(s) qui auraient transmis l'insertion au cultivar R570. Cette incohérence est assez dommageable pour l'interprétation générale puisque la variété POJ 2878 est la variété qui a consacré le succès des programmes d'introggression d'il y a un siècle et qui est la plus impliquée dans la généalogie des cultivars actuels.

Le travail d'analyse de la collection de *Saccharum* comprend également des limites puisque notre échantillon est de taille assez faible et que, comme mentionné précédemment, nous savons qu'il existe de nombreuses erreurs d'« étiquette » dans les collections. Cette analyse nous a montré que cette insertion est très présente chez les « barberi » (90% des clones analysés) et par contre peu fréquente chez les *S. officinarum*, *S. spontaneum* et *S. robustum* (14, 13 et 7 % des clones analysés respectivement). D'autre part, les espèces *S. officinarum*, *S. robustum* et *S. barberi* / *S. sinense* semblent contenir l'ensemble de l'insertion alors que chez *S. spontaneum* les clones présentent des parties variable de l'insertion. Une interprétation de ces données est que la région correspondant à l'insertion peut recombiner dans un contexte *S. spontaneum* mais est fixé dans le contexte génomique des autres espèces. Le fait que cette région puisse recombiner dans un contexte *S. spontaneum* appuierait l'hypothèse que cette région provienne de cette espèce.

Nous avons retrouvé l'insertion chez quelques clones de *S. officinarum* et *S. robustum*. Ce résultat pourrait être dû au fait que certains de ces clones aient été introgressés par *S. spontaneum* et/ou qu'il s'agisse de clones mal classés (erreur d'« étiquettes »). Une étude de la diversité au sein de l'espèce *S. officinarum* et des cultivars modernes a montré, en utilisant des marqueurs RFLP, que certains clones *S. officinarum* n'étaient pas des *officinarum* purs et possédaient probablement des introgressions issues d'autres espèces de *Saccharum*, en particulier avec *S. spontaneum* (Jannoo *et al.*, 1999). Ces auteurs ont utilisé un indice permettant une estimation indirecte de l'hétérozygotie chez les clones : nombre de marqueurs par combinaisons de sonde/enzyme (Nf/c). Cet indice s'est révélé égale à la valeur observée chez les cultivars modernes pour certains clones classés comme *officinarum* et donc les auteurs les ont considérés comme des *officinarum* non purs. Un certain nombre de clones de cette étude sont communs avec notre échantillon. Cinq des *S. officinarum* de notre collection que nous avons montré comme présentant l'insertion avaient été classés comme non purs par Jannoo *et al* (1999). Pour les 5 clones restant, trois ont été analysés par Jannoo *et al*, (1999) et n'ont pas été classés comme non purs. Piperidis *et al* (2010) ont comparé la quantité d'ADN de 255 clones classés comme *S. officinarum* avec un *S. officinarum* témoin à 80 chromosomes



(= nombre caractéristique de cette espèce) : 11 % des clones ont présenté une quantité d'ADN supérieure et donc vraisemblablement un nombre de chromosomes supérieur à 80. Cinq clones appartenant à cette catégorie ont été analysés par la technique GISH et les auteurs ont montré que dans les 5 cas, ces clones étaient des hybrides contenant quelques chromosomes de *S. spontaneum*. Ceci corrobore le fait que les clones de cette espèce présentant un nombre plus élevé de chromosomes sont généralement considérés comme des hybrides (Bremer, 1961; Daniels & Roach, 1987; Sreenivasan *et al.*, 1987).

Les hypothèses actuelles sur les centres d'origines des différents groupes de *Saccharum* suggèrent que *S. officinarum* aurait été domestiquée en Papouasie-Nouvelle-Guinée à partir de *S. robustum*. *S. robustum* est une espèce endémique de la Papouasie-Nouvelle-Guinée. Le centre d'origine de *S. spontaneum* se situerait dans une région entre l'Inde et la Chine mais son aire de répartition est beaucoup plus vaste : du Japon jusqu'à l'Afrique de l'Est. Différentes hypothèses ont été émises quant à l'origine des groupes « barberi » et « sinense ». La plus largement acceptée est celle proposée par Brandes (1956) qui propose que lors de leur migration, les hommes ont transporté des plants de *S. officinarum* vers le nord ouest et l'est. En Inde et Chine, ces clones se seraient hybridés avec *S. spontaneum* et auraient permis l'émergence vers 1500 ans avant notre ère de cultivars mieux adaptés aux conditions subtropicales, les « barberi/sinense ». C'est vraisemblablement en Inde/Chine avec ces clones que l'industrie du sucre est née. Brandes (1956) suggère également qu'au cours de ces migrations ces clones *S. officinarum* ont pu être un peu modifiés par hybridation avec *S. spontaneum*. Piperidis *et al.*, (2010) ont montré par cytogénétique moléculaire que des clones de *S. officinarum* collectés en Indonésie comportés quelques chromosomes de *S. spontaneum*. Il est intéressant de noter que plusieurs groupes de *S. barberi/S. sinense* portent l'insertion porteuse de *Bru1* qui *a priori* est issue d'un événement unique de translocation. Ce résultat suggère que ces groupes ont un ancêtre commun qui leur aurait transmis cette insertion.

Cette introgression pourrait alors provenir d'un *S. officinarum* ou d'un *S. spontaneum* via un clone de *S. officinarum* introgressé par *S. spontaneum* avant la migration vers l'Inde / la Chine ou à l'arrivée dans ces pays. La rouille brune a été signalée pour la première fois en Inde dans les années 1950 quand une épidémie s'est développée sur un plant de canne à sucre cultivé depuis quelques années (Ryan & Egan, 1989). Puis elle a été décrite en 1953 en Chine dans la région du Yunnan (pays frontalier de l'Inde). Dans les années 1970, deux autres pays proches de la Chine sont touchés : le Japon et Taïwan. Puis en 1978, nous avons une épidémie



en Australie et dans la région des Caraïbes. Cette maladie est à ce jour toujours absente de Papouasie-Nouvelle-Guinée où a été domestiquée *S. officinarum*. On peut donc supposer que seuls les clones résistants à la rouille brune ont pu se développer (si la rouille était déjà présente) ou subsister (lorsque la rouille est apparue) en Inde et en Chine.

Les résultats obtenus sur l'origine de l'insertion indiquent que cette insertion est ancienne et suggèrent qu'elle aurait été transmise aux cultivars modernes *via* l'espèce *S. barberi* qui a été utilisée il y a un siècle dans les programmes d'amélioration de la canne à sucre pour rechercher des résistances aux maladies. Concernant l'origine spécifique *S. spontaneum* vs *S. officinarum*, nous ne pouvons actuellement conclure. Le fait que chez *S. spontaneum* on retrouve des portions variables de cette région peut être interprété comme le fait qu'elle recombine au sein de cette espèce et donc en serait dérivée. Par contre, le fait que l'on trouve apparemment cette région dans des *S. officinarum* de Papouasie-Nouvelle-Guinée qui ont *a priori* 80 chromosomes et n'ont pas été soumis à des introgressions avec *S. spontaneum* soutiendrait plutôt une origine *S. officinarum*.

Pour tester ces deux hypothèses, il serait important de vérifier si les clones *S. officinarum* dans lesquels l'insertion est détectée sont bien purs. Ceci pourra être réalisé par

hybridation génomique in situ (GISH) qui permet de différencier les chromosomes *S. officinarum* et *S. spontaneum*. Une étude de diversité pourrait également être effectuée sur des gènes contenus dans l'insertion par amplification, clonage et séquençage chez différents clones de *S. officinarum* et *S. spontaneum* afin de comparer aux allèles détectés dans l'insertion chez R570. D'autre part, il serait intéressant de déterminer si et où se trouve la région homologe à l'insertion chez *S. officinarum* et *S. spontaneum*. Ceci pourrait être réalisé en cartographiant des marqueurs co-dominants (SSR, RFLP, ..), développés à partir de la séquence de l'insertion, par exemple dans le croisement R570 X MQ76-53 (hybride entre un cultivar moderne et un *S. spontaneum*).



## Conclusion générale et perspectives

L'objectif premier de ce travail était de combler la carte physique de l'haplotype cible. Les stratégies utilisées nous ont permis de progresser vers cet objectif et ce malgré la complexité du génome de la canne à sucre. La carte physique partielle de l'haplotype cible ne comporte plus qu'un seul espace non couvert qui se situe au niveau d'une insertion porteuse du gène *Bru1*. La présence de cette insertion nous a empêché de continuer à utiliser les stratégies basées sur la synténie et nous a forcé à utiliser une démarche de « marche » sur le chromosome particulièrement fastidieuse dans un contexte hautement polyploïde. L'augmentation de la couverture du génome en clones BAC s'est révélée être fructueuse.

Différentes pistes sont envisagées pour poursuivre le clonage positionnel du gène *Bru1* :

- La poursuite de la marche sur chromosome avec la production de plus de clones BAC pour combler le dernier espace présent sur l'haplotype cible. Cette technique reste lourde et pour tenter de minimiser les étapes de tri des BACs, un criblage au niveau des transformations générées pour la construction de banque BAC avec des amorces PCR spécifiques de l'haplotype cible pourra être réalisé dans un premier temps. Puis seules les transformations amplifiées seront étalées sur les boîtes de pétri. Cette technique est utilisée au CNRGV (Centre National de Ressources Génétiques Végétales) à Toulouse et a été dérivée de la technique décrite par Isidore *et al.*, (2005).
- Une autre approche, basée sur le séquençage en profondeur de fragments d'ADN de bulks résistants et de bulks sensibles, est en cours de discussion avec une équipe du groupe Syngenta.
- La difficulté d'obtention de marqueurs haplotypes-spécifiques chez la canne à sucre reste une contrainte importante dans un contexte de clonage positionnel. Le développement de marqueurs ISBP qui ciblent les sites d'insertions des éléments transposables (Paux *et al.*, 2006; Paux *et al.*, 2010), initié avec succès au cours de cette thèse, pourra être poursuivi. Ce succès est à relier au fait qu'il a été montré une absence générale de colinéarité des éléments transposables entre haplotypes hom(é)ologues chez la canne à sucre (Garsmeur *et al.*, 2010).
- Le fait que *Bru1* soit localisé dans une région (insertion) qui ne recombine pas





est une difficulté supplémentaire dans ce projet de clonage positionnel. Puisqu'il n'est pas possible actuellement par génétique de réduire la région cible, il serait intéressant de trouver un clone dans lequel cette région est présente dans sa position originale (non transloquée) et donc recombine. Par exemple, l'utilisation d'un des clones de *S. spontaneum* de cette étude, si notre hypothèse d'une origine *S. spontaneum* se vérifie, pourrait être utilisé pour produire une descendance pour essayer de réduire la taille de la région cible.

Enfin, nous avons identifié un gène candidat qui correspond à une sérine / thréonine kinase qui présente une forte homologie de séquence avec le gène *Rpg1* de résistance impliqué dans l'interaction *Puccinia graminis* / orge. L'expression de ce gène n'a pas été détectée en PCR dans des conditions normales (sans inoculation du pathogène) chez la canne à sucre. Pour approfondir cette question, des tests d'expression du gène candidat sous condition d'infection (inoculation du pathogène) pourront être réalisés. D'autre part, il serait intéressant de tester ce gène par transformation d'un cultivar sensible. Ceci est en cours de discussion avec l'équipe Australienne de Robert Birch (Université Queensland, Brisbane) qui est expert dans ces techniques chez la canne à sucre. Les techniques de transformation génétique chez la canne fonctionnent bien mais elles sont associées à de fréquents problèmes de « silencing » chez cette espèce (Manickavasagam *et al.*, 2004; Mudge *et al.*, 2009). Deux stratégies pourraient être utilisées pour intégrer le gène dans une plante sensible : soit une insertion de l'ADN génomique complet mais la taille de 12 kb reste une contrainte, soit l'insertion de l'ADNc si nous arrivons à amplifier son produit d'expression.



---

---

## Références Bibliographiques

---

---



- Adhikari KN, McIntosh RA, Oates JD. 2000.** Distribution and temperature sensitivities of genes for stem rust resistance in Australian oat cultivars and selected germplasm. *Australian Journal of Agricultural Research* **51**: 75-83.
- Aitken K, Jackson P, McIntyre C. 2005.** A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. *Theoretical and Applied Genetics* **110**: 789-801.
- Aljanabi SM, Honeycutt RJ, McClelland M, Sobral BWS. 1993.** A genetic-linkage map of *Saccharum spontaneum* L. *Genetics* **134**: 1249-1260.
- Alpy F, Legueux F, Bianchetti L, Tomasetto C. 2009.** Start domain-containing proteins: A review of their role in lipid transport and exchange. *M S-Medecine Sciences* **25**: 181-191.
- Arceneaux G. 1965.** Cultivated sugarcane of the world and their botanical derivation. *Proc. Int. Soc. of Sugarcane Technol.* **12**: 844-854.
- Asnaghi C, D'Hont A, Glaszmann JC, Rott P. 2001.** Resistance of sugarcane cultivar r 570 to *Puccinia melanocephala* isolates from different geographic locations. *Plant Disease* **85**: 282-286.
- Asnaghi C, Paulet F, Kaye C, Grivet L, Deu M, Glaszmann JC, D'Hont A. 2000.** Application of synteny across Poaceae to determine the map location of a sugarcane rust resistance gene. *Theoretical and Applied Genetics* **101**: 962-969.
- Asnaghi C, Roques D, Ruffel S, Kaye C, Hoarau JY, Telismart H, Girard JC, Raboin LM, Risterucci AM, Grivet L, D'Hont A. 2004.** Targeted mapping of a sugarcane rust resistance gene (*bru1*) using bulked segregant analysis and AFLP markers. *Theoretical and Applied Genetics* **108**: 759-764.
- Ayliffe M, Singh R, Lagudah E. 2008.** Durable resistance to wheat stem rust needed. *Current Opinion in Plant Biology* **11**: 187-192.
- Bennetzen JL. 2005.** Transposable elements, gene creation and genome rearrangement in flowering plants. *Current Opinion in Genetics & Development* **15**: 621-627.
- Blume, H. (1985).** *Geography of Sugarcane*. Bartens, Berlin, pp 75 and 203
- Bhat SR, Gill SS. 1985.** The implications of 2n egg gametes in nobilization and breeding of sugarcane. *Euphytica* **34**: 377-384.
- Boivin K, Deu M, Rami JF, Trouche G, Hamon P. 1999.** Towards a saturated sorghum map using RFLP and AFLP markers. *Theoretical and Applied Genetics* **98**: 320-328.
- Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lenington J, Li ZK, Lin YR, Liu SC, Luo LJ, Marler BS, Ming RG, Mitchell SE, Qiang D, Reischmann K, Schulze SR, Skinner DN, Wang YW, Kresovich S, Schertz KF, Paterson AH. 2003.** A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. *Genetics* **165**: 367-386.
- Brandes EW. 1956.** Origin, dispersal and use in breeding of the Melanesian garden sugarcane and their derivatives, *Saccharum officinarum* L. . *Proceedings of the International Society of Sugar and Sugar Cane Technologists* **9**: 709-750.
- Brandes EW, Sartoris GB, Grassl CO. 1939.** Assembling and evaluating wild forms of sugarcane and closely related plants. *Proc. Int. Soc. of Sugarcane Technol.* **6**: 128-154.
- Bremer G. 1932.** On the somatic chromosome numbers of sugarcane forms and the chromosome numbers of indigenous Indian canes. *Prot Int Soc Sugarcane Technol* **4**: 1-3.
- Bremer G. 1961.** Problems in breeding and cytology of sugar cane. *Euphytica* **10**: 59-78.
- Bremer G. 1966.** Origin of north Indian sugarcane. *Genetica* **37**: 345-363.



- Brueggeman R, Drader T, Kleinhofs A. 2006.** The barley serine/threonine kinase gene rpg1 providing resistance to stem rust belongs to a gene family with five other members encoding kinase domains. *Theoretical and Applied Genetics* **113**: 1147-1158.
- Brueggeman R, Druka A, Nirmala J, Cavileer T, Drader T, Rostoks N, Mirlohi A, Bennypaul H, Gill U, Kudrna D, Whitelaw C, Kilian A, Han F, Sunl Y, Gill K, Steffenson B, Kleinhofs A. 2008.** The stem rust resistance gene rpg5 encodes a protein with nucleotide-binding-site, leucine-rich, and protein kinase domains. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 14970-14975.
- Brueggeman R, Rostoks N, Kudrna D, Kilian A, Han F, Chen J, Druka A, Steffenson B, Kleinhofs A. 2002.** The barley stem rust-resistance gene rpg1 is a novel disease-resistance gene with homology to receptor kinases. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 9328-9333.
- Brunner S, Keller B, Feuillet C. 2003.** A large rearrangement involving genes and low-copy DNA interrupts the microcollinearity between rice and barley at the rph7 locus. *Genetics* **164**: 673-683.
- Chelkowski J, Tyrka M, Sobkiewicz A. 2003.** Resistance genes in barley (*hordeum vulgare* L.) and their identification with molecular markers. *J Appl Genet* **44**: 291-309.
- Chen XM. 2005.** Epidemiology and control of stripe rust [*puccinia striiformis* f. Sp tritici] on wheat. *Canadian Journal of Plant Pathology-Revue Canadienne De Phytopathologie* **27**: 314-337.
- Chen XW, Shang JJ, Chen DX, Lei CL, Zou Y, Zhai WX, Liu GZ, Xu JC, Ling ZZ, Cao G, Ma BT, Wang YP, Zhao XF, Li SG, Zhu LH. 2006.** A b-lectin receptor kinase gene conferring rice blast resistance. *Plant Journal* **46**: 794-804.
- Choi HK, Mun JH, Kim DJ, Zhu HY, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR. 2004.** Estimating genome conservation between crop and model legume species. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 15289-15294.
- Collins N, Drake J, Ayliffe M, Sun Q, Ellis J, Hulbert S, Pryor T. 1999.** Molecular characterization of the maize rp1-d rust resistance haplotype and its mutants. *Plant Cell* **11**: 1365-1376.
- Comstock JC 1985.** A new race of rust suspected. In. Hawai: Hawai Sugar Planters Association Experiment Station.
- Comstock JC, Raid RN 1994.** Sugarcane common rust. In: R. G.P, G. A.G, U. P.P, B. A., A. V.P.C. C.T eds. *Current trends in sugarcane pathology*. Delhi: International Books and Periodicals Supply Service, 1-10.
- Comstock JC, Shine JM, Raid RN. 1992b.** Effect of early rust infection on subsequent sugarcane growth. *Sugar Cane* **4**: 7-9.
- Comstock JC, Sood SG, Glynn NC, Shine JM, McKemy JM, Castlebury LA. 2008.** First report of *puccinia kuehnii*, causal agent of orange rust of sugarcane, in the united states and western hemisphere. *Plant Disease* **92**: 175-175.
- Comstock JC, Wu KK, R.J S. 1992a.** Heritability of resistance to sugar cane rust. *Sugar Cane* **6**: 7-10.
- Conte MG, Gaillard S, Lanau N, Rouard M, Perin C. 2008.** Greenphyldb: A database for plant comparative genomics. *Nucleic Acids Research* **36**: D991-D998.
- Costet L, Le Cunff L, Royeart S, Hoarau J-Y, Raboin LM, Garsmeur O, Pauquet J, Glaszmann JC, Nibouche S, D'Hont A. en préparation.** Ld based tagging of the commun rust (*puccinia melanocephala* h. & p. Sydow) resistance gene brul in sugarcane cultivars.





- Cuadrado A, Acevedo R, de la Espina SMD, Jouve N, de la Torre C. 2004.** Genome remodelling in three modern s-officinarumxs-spontaneum sugarcane cultivars. *Journal of Experimental Botany* **55**: 847-854.
- Cummins GB, Hiratsuka Y. 1983.** *Illustrated genera of rust fungi*. St. Paul: Minn.
- Cummins GB, Hiratsuka Y. 2003.** *Illustrated genera of rust fungi*. St. Paul: Minn.
- D'Hont A, Glaszmann JC. 2001.** *Sugarcane genome analysis with molecular markers: A first decade of research*. Mackay: Australian Soc Sugar Cane Technologists.
- D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann JC. 1996.** Characterisation of the double genome structure of modern sugarcane cultivars (saccharum spp.) by molecular cytogenetics. *Mol Gen Genet* **250**: 405-413.
- D'Hont A, Ison D, Alix K, Roux C, Glaszmann JC. 1998.** Determination of basic chromosome numbers in the genus saccharum by physical mapping of ribosomal rna genes. *Genome* **41**: 221-225.
- D'Hont A, Lu YH, Leon DG, Grivet L, Feldmann P, Lanaud C, Glaszmann JC. 1994.** A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the andropogoneae tribe. *Genome* **37**: 222-230.
- D'Hont A, Paulet F, Glaszmann JC. 2002.** Oligoclonal interspecific origin of 'north indian' and 'chinese' sugarcanes. *Chromosome Research* **10**: 253-262.
- Dangl JL, Jones JDG. 2001.** Plant pathogens and integrated defence responses to infection. *Nature* **411**: 826-833.
- Daniels J, Roach BT 1987.** Taxonomy and evolution. In: H. D.J. ed. *Sugarcane improvement through breeding*. New-York: Elsevier Press, 7-84.
- Daniels J, Roach BT, Daniels C, Panton N. 1991.** The taxonomic status of *saccharum barberi* jeswiet and s. *Sinense* roxb. *Sugarcane* **3**: 11-16.
- Dardick C, Ronald P. 2006.** Plant and animal pathogen recognition receptors signal through non-rd kinases. *Plos Pathogens* **2**: 14-28.
- Daugrois JH, Grivet L, Roques D, Hoarau JY, Lombard H, Glaszmann JC, DHont A. 1996.** A putative major gene for rust resistance linked with a rflp marker in sugarcane cultivar 'r570'. *Theoretical and Applied Genetics* **92**: 1059-1064.
- De Vries SC, Van de Ven GWJ, Van Ittersum MK, Giller KE. 2010.** Resource use efficiency and environmental performance of nine major biofuel crops, processed by first-generation conversion techniques. *Biomass and Bioenergy* **34**: 588 – 601.
- Dean JL, Purdy LH. 1984.** Races of the sugarcane rust fungus, *puccinia melanocephala*, found in florida. *Sugar Cane* **1**: 15-16.
- Deslandes L, Olivier J, Theulieres F, Hirsch J, Feng DX, Bittner-Eddy P, Beynon J, Marco Y. 2002.** Resistance to *ralstonia solanacearum* in *arabidopsis thaliana* is conferred by the recessive *rrs1-r* gene, a member of a novel family of resistance genes. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 2404-2409.
- Develey-Riviere MP, Galiana E. 2007.** Resistance to pathogens and host developmental stage: A multifaceted relationship within the plant kingdom. *New Phytologist* **175**: 405-416.
- Dietel P. 1928.** *Reihe uredinales. In die natürlichen pflanzenfamilien*. Leipzig: Wilhelm Engelmann.
- Dixon LJ, Castlebury LA, Aime C, Glynn NC, Comstock JC. 2010.** Phylogenetic relationships of sugarcane rust fungi. *Mycol Progress* **9**.
- Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008.** Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* **42**: 443-461.



- Dufour P, Deu M, Grivet L, Dhont A, Paulet F, Bouet A, Lanaud C, Glaszmann JC, Hamon P. 1997. Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. *Theoretical and Applied Genetics* **94**: 409-418.
- Feeling M. 2001. Grasses as a single genetic system. Reassessment 2001. *Plant Physiol* **125**: 1191-1197.
- Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B. 2003. Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*triticum aestivum* L.) genome. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 15253-15258.
- Flor HH. 1971. Current status of the gene-for-gene concept. *Annual Review of Phytopathology* **9**: 275-296.
- Flores RC, Loyo JR, Ojeda RA, Rangel OCA, Cer  n FA, M  rquez W, Guerra-Moreno AS, Hernandez-Ibarra HM, Gonz  lez RE, Castlebury LA, Dixon LJ, Glynn NC, Comstock JC, Flynn J, Amador J. 2009. First report of orange rust of sugarcane caused by *puccinia kuehnii* in Mexico, El Salvador, and Panama. *Plant Disease* **93**: 1347-1347.
- Foissac S, Gouzy J, Rombauts S, Mathe C, Amselem J, Sterck L, Van de Peer Y, Rouze P, Schiex T. 2008. Genome annotation in plants and fungi: Eugene as a model platform. *Current Bioinformatics* **3**: 87-97.
- Fu DL, Uauy C, Distelfeld A, Blechl A, Epstein L, Chen XM, Sela HA, Fahima T, Dubcovsky J. 2009. A kinase-start gene confers temperature-dependent resistance to wheat stripe rust. *Science* **323**: 1357-1360.
- Gale MD, Devos KM. 1998. Comparative genetics in the grasses. *Proc Natl Acad Sci U S A* **95**: 1971-1974.
- Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DMA, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**: 498-511.
- Garsmeur O, Charron C, Bocs S, Jouffe V, Samain S, Couloux A, Droc G, Zini C, Glaszmann JC, Van Sluys MA, D'Hont A. 2010. High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane. *New Phytologist*.
- Gill KS, Lubbers EL, Gill BS, Raupp WJ, Cox TS. 1991. A genetic linkage map of *triticum tauschii* (dd) and its relationship to the d genome of bread wheat (aabbdd). *Genome* **34**: 362-374.
- Glaszmann JC, Dufour P, Grivet L, D'Hont A, Deu M, Paulet F, Hamon P. 1997. Comparative genome analysis between several tropical grasses. *Euphytica* **96**: 13-21.
- Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R,



- Feldhaus J, Macalma T, Oliphant A, Briggs S. 2002.** A draft sequence of the rice genome (oryza sativa l. Ssp. Japonica). *Science* **296**: 92-100.
- Grivet L. 1995.** *Marquage moléculaire chez la canne à sucre (saccharum spp.) : Décomposition d'une structure génétique complexe et application à l'amélioration variétale.* Université de Paris XI.
- Grivet L, D'Hont A, Dufour P, Hamon P, Roques D, Glaszmann JC. 1994.** Comparative genome mapping of sugarcane with other species within the andropogoneae tribe. *Heredity* **73**: 500-508.
- Grivet L, D'Hont A, Roques D, Feldmann P, Lanaud C, Glaszmann JC. 1996.** Rflp mapping in cultivated sugarcane (*saccharum* spp.): Genome organization in a highly polyploid and aneuploid interspecific hybrid. *Genetics* **142**: 987-1000.
- Grivet L, Glaszmann JC, D'Hont A. 2006.** Molecular evidence of sugarcane evolution and domestication. In: Motley Timothy J. , Zerega Nyree C. Hugh eds. *Darwin's harvest : New approaches to the origins, evolution, and conservation of crops.* New York Columbia University Press 49-66.
- Guimaraes CT, Honeycutt RJ, Sills GR, Sobral BWS. 1999.** Genetic maps of *saccharum officinarum* l. And *saccharum robustum* brandes & jew. Ex grassl. *Genetics and Molecular Biology* **22**: 125-132.
- Guimaraes CT, Sills GR, Sobral BWS. 1997.** Comparative mapping of andropogoneae: *Saccharum* l. (sugarcane) and its relation to sorghum and maize. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 14261-14266.
- Hammond-Kosack KE, Parker JE. 2003.** Deciphering plant-pathogen communication: Fresh perspectives for molecular resistance breeding. *Current Opinion in Biotechnology* **14**: 177-193.
- Han F, Kilian A, Chen JP, Kudrna D, Steffenson B, Yamamoto K, Matsumoto T, Sasaki T, Kleinhofs A. 1999.** Sequence analysis of a rice bac covering the syntenous barley rpg1 region. *Genome* **42**: 1071-1076.
- Hanks SK, Hunter T. 1995.** Protein kinases .6. The eukaryotic protein-kinase superfamily - kinase (catalytic) domain-structure and classification. *Faseb Journal* **9**: 576-596.
- Harder DE. 1999.** Usefulness of gene pg10 as a source of stem rust resistance in oat breeding. *Phytopathology* **89**: 1214-1217.
- Henrissat B, Coutinho PM, Davies GJ. 2001.** A census of carbohydrate-active enzymes in the genome of *arabidopsis thaliana*. *Plant Molecular Biology* **47**: 55-72.
- Hoarau JY, Offmann B, D'Hont A, Risterucci AM, Roques D, Glaszmann JC, Grivet L. 2001.** Genetic dissection of a modern sugarcane cultivar (*saccharum* spp.). I. Genome mapping with aflp markers. *Theoretical and Applied Genetics* **103**: 84-97.
- Holland JB, Uhr DV, Jeffers D, Goodman MM. 1998.** Inheritance of resistance to southern corn rust in tropical by corn-belt maize populations. *Theoretical and Applied Genetics* **96**: 232-241.
- Horvath H, Rostoks N, Brueggeman R, Steffenson B, von Wettstein D, Kleinhofs A. 2003.** Genetically engineered stem rust resistance in barley using the rpg1 gene. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 364-369.
- Huang L, Brooks SA, Li WL, Fellers JP, Trick HN, Gill BS. 2003.** Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. *Genetics* **164**: 655-664.
- Ilic K, SanMiguel PJ, Bennetzen JL. 2003.** A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 12265-12270.



- Isidore E, Scherrer B, Bellec A, Budin K, Faivre-Rampant P, Waugh R, Keller B, Caboche M, Feuillet C, Chalhoub B. 2005.** Direct targeting and rapid isolation of bac clones spanning a defined chromosome region. *Funct Integr Genomics* **5**: 97-103.
- Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A. 2007.** Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. *Plant Journal* **50**: 574-585.
- Jannoo N, Grivet L, Seguin M, Paulet F, Domaingue R, Rao PS, Dookun A, D'Hont A, Glaszmann JC. 1999.** Molecular investigation of the genetic base of sugarcane cultivars. *Theoretical and Applied Genetics* **99**: 171-184.
- Joosten M, de Wit P. 1999.** The tomato-cladosporium fulvum interaction: A versatile experimental system to study plant-pathogen interactions. *Annual Review of Phytopathology* **37**: 335-367.
- Kang HY, Cox DW. 1996.** Tandem repeats 3' of the igha genes in the human immunoglobulin heavy chain gene cluster. *Genomics* **35**: 189-195.
- Keller B, Feuillet C, Yahiaoui N. 2005.** Map-based isolation of disease resistance genes from bread wheat: Cloning in a supersize genome. *Genetical Research* **85**: 93-100.
- Kilian A, Kudrna DA, Kleinhofs A, Yano M, Kurata N, Steffenson B, Sasaki T. 1995.** Rice-barley synteny and its application to saturation mapping of the barley rpg1 region. *Nucleic Acids Research* **23**: 2729-2733.
- Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL, Keller B. 2009.** A putative abc transporter confers durable resistance to multiple fungal pathogens in wheat. *Science* **323**: 1360-1363.
- Lam E, Shine J, Da Silva J, Lawton M, Bonos S, Calvino M, Carrer H, Silva-Filho MC, Glynn N, Helsel Z, Ma J, Richard E, Souza GM, Ming R. 2009.** Improving sugarcane for biofuel: Engineering for an even better feedstock. *Global Change Biology Bioenergy* **1**: 251-255.
- Le Cunff L, Garsmeur O, Raboin LM, Pauquet J, Telismart H, Selvi A, Grivet L, Philippe R, Begum D, Deu M, Costet L, Wing R, Glaszmann JC, D'Hont A. 2008.** Diploid/polyploid syntenic shuttle mapping and haplotype-specific chromosome walking toward a rust resistance gene (bru1) in highly polyploid sugarcane (2n similar to 12x similar to 115). *Genetics* **180**: 649-660.
- Leonard KJ, Szabo LJ. 2005.** Stem rust of small grains and grasses caused by puccinia graminis. *Molecular Plant Pathology* **6**: 99-111.
- Lepoivre M. 2003.** *Phytopathologie : Bases moléculaires et biologiques des pathosystèmes et fondements des stratégies de lutte*. Bruxelles: De Boeck and Larcier.
- Levy I, Shani Z, Shoseyov O. 2002.** Modification of polysaccharides and plant cell wall by endo-1,4-beta-glucanase and cellulose-binding domains. *Biomolecular Engineering* **19**: 17-30.
- Lu Y, D'Hont A, Paulet F, Grivet L, Arnaud M, Glaszmann JC. 1994a.** Molecular diversity and genome structure in modern sugarcane varieties. *Euphytica* **78**: 217-226.
- Lu Y, D'Hont A, Walker DIT, Rao PS, Feldmann P, Glaszmann JC. 1994b.** Relationships among ancestral species of sugarcane revealed with rflp using single-copy maize nuclear probes. *Euphytica* **78**: 7-18.
- Maccheroni W, Jordao H, Degaspari R, Matsuoka S. 2007.** Development of a dependable microsatellite-based fingerprinting system for sugarcane. *Proc. Int. Soc. of Sugar Cane Technol.* **26**: 889-900.
- Mackey D, Holt BF, Wiig A, Dangl JL. 2002.** Rin4 interacts with pseudomonas syringae type iii effector molecules and is required for rpm1-mediated resistance in arabidopsis. *Cell* **108**: 743-754.





- Maier W, Begerow D, Weiss M, Oberwinkler F. 2003.** Phylogeny of the rust fungi: An approach using nuclear large subunit ribosomal DNA sequences. *Canadian Journal of Botany-Revue Canadienne De Botanique* **81**: 12-23.
- Manickavasagam M, Ganapathi A, Anbazhagan VR, Sudhakar B, Selvaraj N, Vasudevan A, Kasthuriangan S. 2004.** Agrobacterium-mediated genetic transformation and development of herbicide-resistant sugarcane (saccharum species hybrids) using axillary buds. *Plant Cell Reports* **23**: 134-143.
- Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganai MW, Spivey R, Wu TY, Earle ED, Tanksley SD. 1993.** Map-based cloning of a protein-kinase gene conferring disease resistance in tomato. *Science* **262**: 1432-1436.
- Michelmore RW, Paran I, Kesseli RV. 1991.** Identification of markers linked to disease-resistance genes by bulked segregant analysis - a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 9828-9832.
- Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH. 1998.** Detailed alignment of saccharum and sorghum chromosomes: Comparative organization of closely related diploid and polyploid genomes. *Genetics* **150**: 1663-1682.
- Minic Z. 2008.** Physiological roles of plant glycoside hydrolases. *Planta* **227**: 723-740.
- Mudge SR, Osabe K, Casu RE, Bonnett GD, Manners JM, Birch RG. 2009.** Efficient silencing of reporter transgenes coupled to known functional promoters in sugarcane, a highly polyploid crop species. *Planta* **229**: 549-558.
- Münch-Garhoff S, Neuhaus J-M, Boller T, Kemmerling B, Kogel K-H. 1997.** Expression of  $\beta$ -1,3-glucanase and chitinase in healthy, stem-rust-affected and elicitor-treated near-isogenic wheat lines showing sr5-or sr24-specified race-specific rust resistance. *Planta* **201**: 235-244.
- Panjabi P, Jagannath A, Bisht NC, Padmaja KL, Sharma S, Gupta V, Pradhan AK, Pental D. 2008.** Comparative mapping of brassica juncea and arabidopsis thaliana using intron polymorphism (ip) markers: Homoeologous relationships, diversification and evolution of the a, b and c brassica genomes. *Bmc Genomics* **9**: 19.
- Panje, R.R. & C.N. Babu. 1960.** Studies in *Saccharum spontaneum*; distribution and geographical association of chromosome numbers. *Cytologia* **25**:152-172.
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Ojillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS. 2009.** The sorghum bicolor genome and the diversification of grasses. *Nature* **457**: 551-556.
- Paton N, Daniels J, Smith P. 1978.** A study of *s. Sinense* roxb. Ans *s. Barberi* jesw. . *Sugarcane Breed Newslett* **41**: 33-50.
- Paux E, Faure S, Choulet F, Roger D, Gauthier V, Martinant JP, Sourdille P, Balfourier F, Le Paslier MC, Chauveau A, Cakir M, Gandon B, Feuillet C. 2010.** Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. *Plant Biotechnology Journal* **8**: 196-210.



- Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P, Feuillet C. 2006.** Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through bac-end sequencing on chromosome 3b. *Plant Journal* **48**: 463-474.
- Piegu B, Guyot R, Picault N, Roulin A, Saniyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O. 2006.** Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in *oryza australiensis*, a wild relative of rice. *Genome Research* **16**: 1262-1269.
- Piperidis G, Piperidis N, D'Hont A. 2010.** Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Mol Genet Genomics* **284**: 65-73.
- Price S. 1957.** Cytological studies in *saccharum* and allied genera. Iii. Chromosome numbers in interspecific hybrids. *Botanical Gazette* **118**: 146-159.
- Price S. 1965.** *Cytology of saccharum robustum and related sympatric species and natural hybrids*. Washington, DC: Bull . 1337 . U .S . Gov . Print . Office.
- Price S. 1968.** Cytology of chinese and north indian sugarcane. *Econ Botany* **22**: 154-155.
- Purdy LH, Liu LJ, Dean JL. 1983.** Sugarcane rust, a newly important disease. . *Plant Disease* **67**: 1292-1296.
- Qiu JW, Schurch AC, Yahiaoui N, Dong LL, Fan HJ, Zhang ZJ, Keller B, Ling HQ. 2007.** Physical mapping and identification of a candidate for the leaf rust resistance gene *Lr1* of wheat. *Theoretical and Applied Genetics* **115**: 159-168.
- Raboin LM, Oliveira KM, Lecunff L, Telismart H, Roques D, Butterfield M, Hoarau JY, D'Hont A. 2006.** Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: Identification of a gene controlling stalk colour and a new rust resistance gene. *Theoretical and Applied Genetics* **112**: 1382-1391.
- Rafalski JA, Tingey SV. 1993.** Genetic diagnostics in plant-breeding - rapds, microsatellites and machines. *TRENDS in genetics* **9**: 275-280.
- Ravin NV, Ravin VK. 1999.** Use of a linear multicopy vector based on the mini-replicon of temperate coliphage  $\lambda$ 15 for cloning DNA with abnormal secondary structures. *Nucleic Acids Res* **27**: e13.
- Razin SV, Ioudinkova ES, Trifonov EN, Scherrer K. 2001.** Non-clonability correlates with genomic instability: A case study of a unique DNA region. *Journal of Molecular Biology* **307**: 481-486.
- Roach BT. 1986.** Evaluation and use of sugarcane germplasm *Proceedings of the 19th Congress of the International Society of Sugar Cane Technologists* **1**: 492-503.
- Rossi M, Araujo PG, Paulet F, Garsmeur O, Dias VM, Chen H, Van Sluys MA, D'Hont A. 2003.** Genomic distribution and characterization of est-derived resistance gene analogs (rgas) in sugarcane. *Molecular Genetics and Genomics* **269**: 406-419.
- Rostoks N, Steffenson BJ, Kleinhofs A. 2004.** Structure and expression of the barley stem rust resistance gene *rpg1* messenger rna. *Physiological and Molecular Plant Pathology* **64**: 91-101.
- Rouard M, Guignon V, Aluome C, Laporte M-A, Droc G, Walde C, Zmasek CM, Périn C, Conte MG. 2010.** Greenphyldb v2.0: Comparative and functional genomics in plants *Nucleic Acids Research*: 1-8.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000.** Artemis: Sequence visualization and annotation. *Bioinformatics* **16**: 944-945.
- Ryan CC, Egan BT 1989.** Rust. In: C. Ricaud, B. T. Egan, A. G. GillaspieC. G. Hughes eds. *Disease of sugarcane*. Amsterdam: Elsevier, 189-210.



- Saito K, Hayano-Saito Y, Kuroki M, Sato Y. 2010. Map-based cloning of the rice cold tolerance gene *ctb1*. *Plant Science* 79: 97-102.
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei FS, Pasternak S, Liang CZ, Zhang JW, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du FY, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen WZ, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He RF, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin JK, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren LY, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddelloh JA, Han YJ, Lee H, Li PH, Lisch DR, Liu SZ, Liu ZJ, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang LX, Yu Y, Zhang LF, Zhou SG, Zhu Q, Bennetzen JL, Dawe RK, Jiang JM, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK. 2009. The b73 maize genome: Complexity, diversity, and dynamics. *Science* 326: 1112-1115.
- Shine J, Comstock JC, Dean JL. 2005. Comparison of five isolates of sugarcane rust and differential reaction on six sugarcane clones. *Sugar Cane International* 23: 24-29.
- Silva J, Honeycutt RJ, Burnquist W, Al-Janabi SM, Sorrells ME, Tanksley SD, Sobral BWS. 1995. *Saccharum spontaneum* l. 'ses 208' genetic linkage map combining rflp- and pcr-based markers. *Molecular Breeding* 1: 165-179.
- Silva JA, Sorrells ME, Burnquist WL, Tanksley SD. 1993. Rflp linkage map and genome analysis of *saccharum spontaneum*. *Genome* 36: 782-791.
- Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *xa21*. *Science* 270: 1804-1806.
- Sreenivasan TV, Ahloowalia BS, Heinz DJ 1987. Cytogenetics. In: H. DJ ed. *Sugarcane improvement through breeding*. New-York: Elsevier, 211-253.
- Srinivasan KV, Muthaiyan MC 1965. A note on physiologic races in *puccinia erianthi* padw. And khan affecting sugar-cane varieties. In Elsevier. *International society of sugar cane technologists congress*. Puerto Rico. 1126-1129.
- Staples RC. 2003. A novel gene for rust resistance. *Trends in Plant Science* 8: 149-151.
- Staskawicz BJ, Ausubel FM, Baker BJ, Ellis JG, Jones JDG. 1995. Molecular-genetics of plant-disease resistance. *Science* 268: 661-667.
- Steffenson BJ. 1992. Analysis of durable resistance to stem rust in barley. *Euphytica* 63: 153-167.
- Stevenson GC. 1965. *Genetics and breeding of sugarcane*. London: Longmans.
- Swann EC, Frieders EM, McLaughlin D. 2001. *Urediniomycetes. In the mycota : Systematics and evolution*. Berlin: Springer Verlag.
- Tai PYP, Miller JD, Dean JL. 1981. Inheritance of resistance to rust in sugarcane. *Field Crops Research* 4: 261-268.



- Tameling WIL, Vossen JH, Albrecht M, Lengauer T, Berden JA, Haring MA, Cornelissen BJC, Takken FLW. 2006.** Mutations in the nb-arc domain of i-2 that impair atp hydrolysis cause autoactivation. *Plant Physiology* **140**: 1233-1245.
- Taylor PW, Hogarth DM, Ryan CC 1982.** Progress report inheritance of resistance to rust disease caused by *puccinia melanocephala*. In. Queensland, Australia: Bureau of sugarcane experiment stations.
- Taylor PWJ. 1992.** Evidence for the existence of a single race of commun rust caused by *puccinia melanocephala*, in australian sugar-cane cultivars. *Australian Journal of Agricultural Research* **43**: 443-450.
- Tomkins JP, Yu Y, Miller-Smith H, Frisch DA, Woo SS, Wing RA. 1999.** A bacterial artificial chromosome library for sugarcane. *Theoretical and Applied Genetics* **99**: 419-424.
- Ueda T, Sato T, Hidema J, Hirouchi T, Yamamoto K, Kumagai T, Yano M. 2005.** Quvr-10, a major quantitative trait locus for ultraviolet-b resistance in rice, encodes cyclobutane pyrimidine dimer photolyase. *Genetics* **171**: 1941-1950.
- Van der Merwe M, Ericson L, Walker J, Thrall PH, Burdon JJ. 2006.** Evolutionary relationships among species of puccinia and uromyces (pucciniaceae, uredinales) inferred from partial protein coding gene phylogenies. *Mycological Research* **2007**: 163-175.
- van Loon LC, Rep M, Pieterse CMJ. 2006.** Significance of inducible defense-related proteins in infected plants. *Annual Review of Phytopathology* **44**: 135-162.
- van Ooijen G, van den Burg HA, Cornelissen BJC, Takken FLW. 2007.** Structure and function of resistance proteins in solanaceous plants. *Annual Review of Phytopathology* **45**: 43-72.
- Varghese JN, Garrett TPJ, Colman PM, Chen L, Hoj PB, Fincher GB. 1994.** 3-dimensional structures of 2 plant beta-glucan endohydrolases with distinct substrate specificities. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 2785-2789.
- Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MIT, Henrique-Silva F, Giglioti EA, Lemos MVE, Coutinho LL, Nobrega MP, Carrer H, Franca SC, Bacci M, Goldman MHS, Gomes SL, Nunes LR, Camargo LEA, Siqueira WJ, Van Sluys MA, Thiemann OH, Kuramae EE, Santelli RV, Marino CL, Targon M, Ferro JA, Silveira HCS, Marini DC, Lemos EGM, Monteiro-Vitorello CB, Tambor JHM, Carraro DM, Roberto PG, Martins VG, Goldman GH, de Oliveira RC, Truffi D, Colombo CA, Rossi M, de Araujo PG, Sculaccio SA, Angella A, Lima MMA, de Rosa VE, Siviero F, Coscrato VE, Machado MA, Grivet L, Di Mauro SMZ, Nobrega FG, Menck CFM, Braga MDV, Telles GP, Cara FAA, Pedrosa G, Meidanis J, Arruda P. 2003.** Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Research* **13**: 2725-2735.
- Vilarinhos AD, Piffanelli P, Lagoda P, Thibivilliers S, Sabau X, Carreel F, D'Hont A. 2003.** Construction and characterization of a bacterial artificial chromosome library of banana (*musa acuminata* colla). *Theoretical and Applied Genetics* **106**: 1102-1106.
- Virtudazo E, Nojima H, Kakishima M. 2001.** Taxonomy of *puccinia* species causing rust diseases on sugarcane. *Mycoscience* **42**: 167-175.
- Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K, Lucas S, Harmon-Smith M, Lail K, Tice H, Grimwood J, McKenzie N, Huo NX, Gu YQ, Lazo GR, Anderson OD, You FM, Luo MC, Dvorak J, Wright J, Febrer M, Idziak D, Hasterok R, Lindquist E, Wang M, Fox SE, Priest HD, Filichkin SA,**





- Givan SA, Bryant DW, Chang JH, Wu HY, Wu W, Hsia AP, Schnable PS, Kalyanaraman A, Barbazuk B, Michael TP, Hazen SP, Bragg JN, Laudencia-Chingcuanco D, Weng YQ, Haberer G, Spannagl M, Mayer K, Rattei T, Mitros T, Lee SJ, Rose JKC, Mueller LA, York TL, Wicker T, Buchmann JP, Tanskanen J, Schulman AH, Gundlach H, de Oliveira AC, Maia LD, Belknap W, Jiang N, Lai JS, Zhu LC, Ma JX, Sun C, Pritham E, Salse J, Murat F, Abrouk M, Bruggmann R, Messing J, Fahlgren N, Sullivan CM, Carrington JC, Chapman EJ, May GD, Zhai JX, Ganssmann M, Gurazada SGR, German M, Meyers BC, Green PJ, Tyler L, Wu JJ, Thomson J, Chen S, Scheller HV, Harholt J, Ulvskov P, Kimbrel JA, Bartley LE, Cao PJ, Jung KH, Sharma MK, Vega-Sanchez M, Ronald P, Dardick CD, De Bodt S, Verelst W, Inze D, Heese M, Schnittger A, Yang XH, Kalluri UC, Tuskan GA, Hua ZH, Vierstra RD, Cui Y, Ouyang SH, Sun QX, Liu ZY, Yilmaz A, Grotewold E, Sibout R, Hematy K, Mouille G, Hofte H, Pelloux J, O'Connor D, Schnable J, Rowe S, Harmon F, Cass CL, Sedbrook JC, Byrne ME, Walsh S, Higgins J, Li PH, Brutnell T, Unver T, Budak H, Belcram H, Charles M, Chalhoub B, Baxter I, *Int Brachypodium I*. 2010. Genome sequencing and analysis of the model grass brachypodium distachyon *Nature* **463**: 763-768.
- Waclawovsky AJ, Sato PM, Lembke CG, Moore PH, Souza GM. 2010. Sugarcane for bioenergy production: An assessment of yield and regulation of sucrose content. *Plant Biotechnology Journal* **8**: 1-14.
- Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, Tanksley SD. 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments. *TAG Theoretical and Applied Genetics* **83**: 294-300.
- Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ. 2006. Sub1a is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* **442**: 705-708.
- Xu WG, Li CX, Hu L, Zhang L, Zhang JZ, Dong HB, Wang GS. 2010. Molecular mapping of powdery mildew resistance gene pmhmk in winter wheat (*triticum aestivum* L.) cultivar zhoumai 22 *Molecular Breeding* **26**: 31-38.
- Yahiaoui N, Srichumpa P, Dudler R, Keller B. 2004. Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene pm3b from hexaploid wheat. *Plant Journal* **37**: 528-538.
- Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. 2003. Positional cloning of the wheat vernalization gene vrn1. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 6263-6268.
- Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang ZX, Kono I, Kurata N, Yano M, Iwata N, Sasaki T. 1998. Expression of xa1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 1663-1668.
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Li J, Liu Z, Qi Q, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Zhao W, Li P, Chen W, Zhang Y, Hu J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Tao M, Zhu L, Yuan



- L, Yang H. 2002.** A draft sequence of the rice genome (oryza sativa l. Ssp. Indica). *Science* **296**: 79-92.
- Zhang SQ, Klessig DF. 2001.** Mapk cascades in plant defense signaling. *Trends in Plant Science* **6**: 520-527.
- Zhou CJ, Chen CX, Cao PX, Wu SW, Sun JW, Jin DM, Wang B. 2007.** Characterization and fine mapping of rppq, a resistance gene to southern corn rust in maize. *Molecular Genetics and Genomics* **278**: 723-728.



---

---

## **Annexe 1.**

---

---



# High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane

Olivier Garsmeur<sup>1\*</sup>, Carine Charron<sup>1\*</sup>, Stéphanie Bocs<sup>1</sup>, Vincent Jouffe<sup>1</sup>, Sylvie Samain<sup>2</sup>, Arnaud Couloux<sup>2</sup>, Gaëtan Droc<sup>1</sup>, Cyrille Zini<sup>1</sup>, Jean-Christophe Glaszmann<sup>1</sup>, Marie-Anne Van Sluys<sup>3</sup> and Angélique D'Hont<sup>1</sup>

<sup>1</sup>CIRAD, UMR 1098 DAP, Avenue Agropolis, TA A96/03, F-34398 Montpellier Cedex 5, France; <sup>2</sup>CEA: Institut de génomique – GENOSCOPE, 2, rue Gaston Crémieux, CP 5706, 91057, Evry Cedex, France; <sup>3</sup>GaTE Lab, Departamento de Botanica, Universidade de São Paulo, Rua do Matão 277, São Paulo, 05508-090 SP, Brazil

## Summary

Author for correspondence:

Angélique D'Hont

Tel: +33 4 67 61 59 27

Email: dhont@cirad.fr

Received: 19 July 2010

Accepted: 2 September 2010

New Phytologist (2010)

doi: 10.1111/j.1469-8137.2010.03497.x

**Key words:** autopolyploid, bacterial artificial chromosome (BAC) sequence, homologous genes, selective constraints, sugarcane, transposable elements.

- Modern sugarcane (*Saccharum* spp.) is the leading sugar crop and a primary energy crop. It has the highest level of 'vertical' redundancy ( $2n = 12x = 120$ ) of all polyploid plants studied to date. It was produced about a century ago through hybridization between two autopolyploid species, namely *S. officinarum* and *S. spontaneum*.
- In order to investigate the genome dynamics in this highly polyploid context, we sequenced and compared seven hom(oe)ologous haplotypes (bacterial artificial chromosome clones).
- Our analysis revealed a high level of gene retention and colinearity, as well as high gene structure and sequence conservation, with an average sequence divergence of 4% for exons. Remarkably, all of the hom(oe)ologous genes were predicted as being functional (except for one gene fragment) and showed signs of evolving under purifying selection, with the exception of genes within segmental duplications. By contrast, transposable elements displayed a general absence of colinearity among hom(oe)ologous haplotypes and appeared to have undergone dynamic expansion in *Saccharum*, compared with sorghum, its close relative in the Andropogonea tribe.
- These results reinforce the general trend emerging from recent studies indicating the diverse and nuanced effect of polyploidy on genome dynamics.

## Introduction

Polyploidy (whole genome duplication) is recognized as a major driving force in plant genome evolution. Its footprint is apparent in the genome of all diploid plants that have been fully sequenced (Jaillon *et al.*, 2007; Fawcett *et al.*, 2009; Salse *et al.*, 2009; Schmutz *et al.*, 2010). Analyses of recent polyploids have shown that polyploidy can have profound genetic and epigenetic consequences, including chromosome rearrangements, gene loss, transposable element (TE) activation, DNA methylation and gene expression modifications (Doyle *et al.*, 2008; Jackson & Chen, 2009). Most of these studies have focused on allopolyploids, whereas the impact of autopolyploidy on genome evolution has received very little attention. Autopolyploids are tradi-

tionally considered to arise within a single species through the doubling of structurally similar homologous genomes (e.g. AA becoming AAAA), whereas allopolyploids arise via interspecific hybridization (e.g. yielding an AB hybrid) and subsequent doubling of nonhomologous (i.e. homoeologous) genomes (AABB). A major distinction between auto- and allopolyploids is their mode of chromosome inheritance as a result of meiotic pairing. Typically, autopolyploids have more than two sets of homologous chromosomes that randomly pair, resulting in polysomic inheritance. In allopolyploids, homologous chromosomes pair together (and not with their homoeolog), resulting in disomic inheritance (Mather, 1936; De Silva *et al.*, 2005). The very limited available data on the consequences of autopolyploidization concern gene expression and suggest less modification than after allopolyploidization during the first generations following genome doubling (Albertin *et al.*,

\*These authors contributed equally to this work.

2005; Martelotto *et al.*, 2005; Lu *et al.*, 2006; Stupar *et al.*, 2007; Church & Spaulding, 2009; Parisod *et al.*, 2010). However, little is currently known about the structural consequences and longer term evolution of genomes in autopolyploids relative to allopolyploids.

Sugarcane is one of the world's most efficient crops in converting solar energy into chemical energy, which makes it the leading sugar crop and a primary energy crop. The domesticated sugar-producing species *Saccharum officinarum* is auto-octoploid ( $2n = 8x = 80$ ) and has been domesticated from the wild autopolyploid species *S. robustum* (mainly  $2n = 60, 80$ ) (Brandes, 1956; Sreenivasan *et al.*, 1987; Grivet *et al.*, 2006). A century ago, breeders incorporated chromosomes from a wild relative species, that is *S. spontaneum*, by hybridization and backcrossing. This wild species is also considered to be of autopolyploid origin, with chromosome numbers ranging from  $2n = 5x = 40$  to  $16x = 128$ , with many aneuploid forms (Panje & Babu, 1960; Sreenivasan *et al.*, 1987). The resulting modern cultivars have a genome size of around 10 Gb and typically have around 120 chromosomes, 70–80% of which are entirely derived from *S. officinarum*, 10–20% from *S. spontaneum* and a few from interspecific recombinations (D'Hont *et al.*, 1996; Cuadrado *et al.*, 2004; D'Hont, 2005; Piperidis *et al.*, 2010). The meiosis of modern sugarcane cultivars mainly involves bivalent pairing (Bremer, 1922; Price, 1963; Burner & Legendre, 1994) and chromosome assortment results from general polysomy with some cases of preferential pairing (Grivet *et al.*, 1996; Hoarau *et al.*, 2001; Jannoo *et al.*, 2004).

Taxonomically, sugarcane belongs to *Saccharum*, which encompasses only polyploid species. Synteny is particularly well conserved between sugarcane and sorghum, its close relative in the Andropogoneae tribe (Grivet *et al.*, 1994; Dufour *et al.*, 1997; Guimaraes *et al.*, 1997; Ming *et al.*, 1998; Le Cunff *et al.*, 2008), from which it diverged 8–10 million yr ago (Mya) (Jannoo *et al.*, 2007). The monoloid genome size of sugarcane, that is 760–930 Mb (D'Hont & Glaszmann, 2001), is close to that of sorghum, that is 730 Mb (Paterson *et al.*, 2009).

The general objective of this study was to investigate genome dynamics, in particular the fate of duplicated genes and TE content, in a highly polyploid context such as sugarcane. Previously, we have analyzed two homoeologous sequences (97 kb and 126 kb) co-existing in the modern sugarcane cultivar R570 (Jannoo *et al.*, 2007). These sequences correspond to the region encompassing the *Adh1* gene, which has already been studied in several Poaceae species (Ilic *et al.*, 2003). Our findings revealed perfect colinearity at the gene level between the two sugarcane homoeologous haplotypes, as well as high gene structure conservation. Apart from the insertion of a few retrotransposable elements, high homology was also observed along the nontranscribed regions. This first analysis of sugarcane haplotype organization at the sequence level suggested that the

high ploidy in sugarcane did not induce generalized reshaping of its genome, thus challenging the idea that polyploidy quickly induces generalized rearrangement of genomes.

The specific objectives of this study were to verify the high level of functional gene retention and colinearity and the absence of TE colinearity in a second region of the genome of R570 and across a larger number of haplotypes, including homologous haplotypes (as opposed to homoeologous haplotypes). In addition, we refined the analysis by studying the selection pressures under which the redundant genes were evolving. We took advantage of a physical map developed in the context of a map-based cloning approach (Le Cunff *et al.*, 2008). Bacterial artificial chromosome (BAC) clones belonging to seven hom(oe)ologous haplotypes were sequenced, enabling a comparison between homoeologous (from *S. spontaneum* and *S. officinarum*) and also between homologous (from *S. spontaneum* or *S. officinarum*) haplotypes.

## Materials and Methods

### BAC clone selection and sequencing

Eight sugarcane BAC clones belonging to seven hom(oe)ologous haplotypes of a region bearing the rust resistance gene *Bru1* and three sorghum BAC clones corresponding to the orthologous sorghum region were identified, as described in Le Cunff *et al.* (2008). The sugarcane BAC clones originated from a R570 sugarcane cultivar BAC library developed by Tomkins *et al.* (1999) (BAC clones named Sh) and from a library built by Le Cunff *et al.* (2008) with DNA of selfed progenies of R570 (BAC clones named ShCIR). The sorghum BAC clones were selected from the SB\_BBc sorghum BAC library constructed with DNA from the bicolor Btx623 genotype by D. Begum *et al.*, unpublished (Clemson University Genomics Institute, Clemson, USA, <http://www.genome.clemson.edu/cgi-bin/orders?&page=productGroup&service=bacrc&productGroup=26>). BAC clones were subcloned and sequenced with an average coverage of 14X using the Sanger method and ABI3730 (Applied Biosystems by Life Technologies, Carlsbad, California, USA) sequencers at Genoscope, France (<http://www.cns.fr/spip/>), and sequences were assembled using Phred/Phrap/consed (<http://www.phrap.com/>). Sequences were submitted to the EMBL sequence database under the following accession numbers (BAC clone names in parentheses): FN431669 (ShCIR9O20), FN431661 (ShCIR12E03), FN431668 (Sh253G12), FN431666 (Sh142J21), FN431664 (Sh53A11), FN431665 (Sh135P16), FN431663 (Sh15N23), FN431667 (Sh197G04) and FN431662 (SB\_BBc24P17c).

### Gene annotation

The gene finding software 'EuGène' (Schiex *et al.*, 2001) was used to perform automatic prediction of the gene



structure on the BAC sequences. The learning dataset of genes used to train 'EuGène' for its predictions was composed of 300 validated rice genes. Several sources of information were integrated in the prediction process, such as similarities with the sugarcane expressed sequence tags (ESTs) of the Gene Index database comprising 282 683 sequences ([http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=s\\_officinarum](http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=s_officinarum)). The polypeptide function was also predicted by integrating several pieces of evidence, such as protein similarities (NCBI-BLASTp on UniProt) (Consortium, 2009) and protein domains (InterProScan on InterPro) (Hunter *et al.*, 2009). The sequence of the rice orthologous segment was recovered from the Rice Genome Annotation Project Database (<http://rice.plantbiology.msu.edu/>) and re-annotated. The manual expertise of predictions was performed using the Artemis annotation tool (Rutherford *et al.*, 2000). All genes were checked and ambiguities were manually corrected.

### TE annotation and classification

TEs were annotated using the REPET package containing two pipelines: TEdenovo and TEannot (<http://urgi.versailles.inra.fr/index.php/urgi/Tools/REPET>). Repeat sequences were first predicted with TEdenovo by an all-by-all comparison of BAC sequences, with the genes already annotated being masked. Repeat sequences, showing at least 80% identity on 80% of their length, were clustered and a consensus sequence was derived from each cluster. These consensus sequences were compared with the Repbase database (Jurka *et al.*, 2005) and classified according to Wicker *et al.* (2007) in class, order, superfamily and family of TEs. In addition, the LTR\_STRUC program (McCarthy & McDonald, 2003) was used to identify complete long terminal repeat (LTR) retrotransposons. LTR structures flanking retrotransposons were checked using dotplot analysis of each BAC sequence against itself. Redundancy between LTR\_STRUC and TEdenovo was removed to obtain a nonredundant library of classified elements. Finally, this library was used with the TEannot pipeline to detect additional TE fragments. All large TEs (all TEs except miniature inverted-repeat TEs, MITEs) were then manually checked with the Artemis annotation tool (Rutherford *et al.*, 2000).

### Sequence analyses

All annotated genes were clustered using the NCBI-BLASTclust program to identify homologous and/or orthologous groups of genes (for rice and sorghum). Sequence alignments of hom(oe)ologous and orthologous genes were performed with ClustalW in the GENALYS (<http://software.cng.fr>) and BioEdit (Hall, 1999) interfaces, and manually corrected. The Nei–Gojabori method

implemented in MEGA4 software (Kumar *et al.*, 1994), with the Juke–Cantor correction, was used to estimate synonymous ( $K_s$ ) and nonsynonymous ( $K_a$ ) substitution rates and  $K_a : K_s$  ratios. Positive or negative selections were assessed by testing hypotheses  $H_1$  ( $K_a > K_s$ ) or  $H_2$  ( $K_a < K_s$ ) against the null hypothesis of no selection ( $H_0$ :  $K_a = K_s$ ), using the  $Z$ -test for selection (Nei & Kumar, 2000) implemented in MEGA; the standard error was calculated by the bootstrap method using 1000 replicates.  $P < 0.05$  was considered to be statistically significant. Genetic distances between sequences and phylogenetic trees were established with Kimura's two-parameter model and the neighbor-joining method, respectively. Divergence and duplication times ( $T$ ) were estimated using  $T = K_s/2k$ , where  $K_s$  is the estimated number of synonymous substitutions per site between homologous sequences and  $k$  is the average synonymous substitution rate, by assuming a rate of  $6.5 \times 10^{-9}$  mutations per synonymous site per year, as reported by Gaut *et al.* (1996) for *Adh* loci of grasses. However, it is important to note that current molecular dating approaches are limited as many variables influence the molecular evolutionary rate (Doyle & Egan, 2009).

The insertion times of LTR retrotransposons were estimated for complete elements with both LTRs and target site duplications (Table 1). LTR sequences were aligned using ClustalW and edited manually if needed. MEGA4 software was used to calculate the rate of substitutions. Insertion dates were estimated using Kimura's two-parameter model, with a mutation rate of  $1.3 \times 10^{-8}$  substitutions per synonymous site per year, as described by Ma & Bennetzen (2004).

Haplotype networks were built with NETWORK v4.5 software (<http://www.fluxus-technology.com/sharenet.htm>) from coding sequences of hom(oe)ologous genes, using the median-joining method, with an equal weight for all sites. Both single nucleotide polymorphisms (SNPs) and insertions–deletions (INDELs) were considered. Phylogenetic trees, haplotype networks and divergence time calculations were used to schematize a haplotype differentiation pattern

**Table 1** Summary of bacterial artificial chromosome (BAC) content

Total number of sugarcane BACs analyzed	8
Combined BAC length	960 kb
Transposable elements	337 kb (35%)
Long terminal repeat (LTR) retrotransposons	260 kb (27%)
Non-LTR retrotransposons	48 kb (5%)
Transposons	29 kb (3%)
Genes	207 kb (22%)
Intergenic regions	416 kb (43%)
Gene density	1 gene per 9 kb
Number of colinear genes	14
Number of genes that show colinearity with sorghum	10
Number of genes that show colinearity with rice	8

(Fig. 4). Comparisons were performed when at least four alleles could be compared and excluding gene fragment 10.

## Results

### Sequencing and annotation of seven hom(oe)ologous regions within the R570 sugarcane cultivar

Eight BAC clones representing seven haplotypes were sequenced and annotated, representing nearly 1 Mb of sugarcane DNA sequence. Fifteen genes with their allelic versions distributed in six or seven haplotypes and 66 large TEs were identified (Fig. 1, Table 1). The region analyzed corresponded to a gene-rich region with an average of one gene every 9 kb. Each hom(oe)ologous haplotype was represented by one BAC clone, except for haplotype VII, which was represented by two BAC clones genetically mapped on the same chromosome but separated on the current physical map by a gap of unknown size. All seven hom(oe)ologous regions partially overlapped, except for haplotypes IV and VI. We could therefore not rule out that these two BAC clones belonged to the same chromosome. The overlap between two hom(oe)ologous regions (excluding haplotype IV and VI) ranged from 9.6 to 73 kb, representing between two and seven genes.

### Comparison of gene content and structure within sugarcane haplotypes and with sorghum and rice orthologous regions

Fifteen genes were present in at least two hom(oe)ologous regions in sugarcane or orthologous regions in rice and sorghum. For all of these genes (except gene 10), the coding sequence could be translated into a complete protein sequence, and thus they were all predicted to be functional. In all hom(oe)ologous versions of gene 10, three exons were missing, and so this gene is referred to as a gene fragment. All hom(oe)ologous alleles of the 14 complete genes showed a conserved exon/intron structure. From the sugarcane TIGR database, highly homologous ESTs covering the full length of six of the genes were identified, whereas only partial coverage was found for the other genes (data not shown). Proteins from other species presenting high homology with the amino acid sequences deduced from the 15 genes were found in the database. The function of 11 could be tentatively attributed (Table 2).

As the overlaps varied between hom(oe)ologous regions, the sugarcane hom(oe)ologous regions could be compared on the basis of a variable number of genes, ranging from two to seven, with an average of four. On this basis, the colinearity and transcriptional orientation were highly conserved between genes 1 and 14 (Fig. 1). A notable exception to strict colinearity involved the presence of two additional predictions corresponding to gene fragment 10 and a dupli-

cation of gene 11 on haplotypes II, V, VI and VII. Gene fragment 10 corresponded to a fragment of a serine/threonine kinase. Gene 11 corresponded to a sorbitol-6-phosphate dehydrogenase (S6PDH).

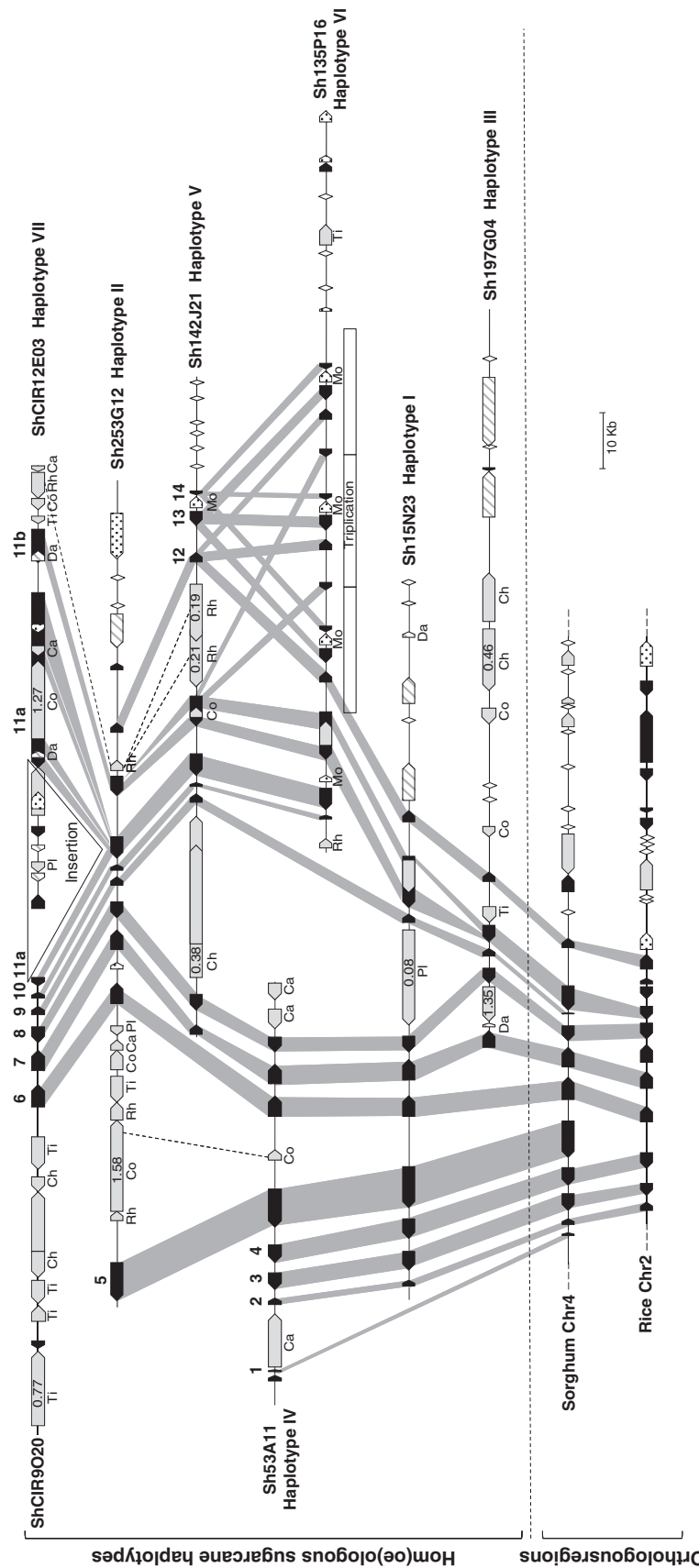
In addition, two structural arrangements were each unique to one haplotype: haplotype VII displayed additional genes between genes 11a and 11b, as the result of an insertion within this haplotype (Le Cunff *et al.*, 2008), and haplotype VI displayed a segmental triplication of 22 kb, including part of gene 11b and genes 12, 13 and 14 (Fig. 1). Finally, a few additional gene predictions were unique for one haplotype, but none was found to interrupt the gene 1–14 series (Fig. 1). Several genes coding for proteinase inhibitor subfamily I13 were identified on five haplotypes (I, II, III, V and VI), thus possibly extending the colinear region.

Colinearity was also very high in the overlapping regions between sugarcane and sorghum and, to a lesser degree, rice. Three genes (genes 9, 13 and 14) were absent in sorghum when compared with sugarcane, but a trace of the first exon still remained for gene 9. In the rice orthologous region, five genes were absent when compared with sugarcane (genes 1, 5, 9, 13 and 14), whereas one additional gene was present between genes 7 and 8 and two were present between genes 11 and 12. As for sugarcane, several proteinase inhibitors of subfamily I13 were identified at a similar position in the sorghum and rice orthologous regions.

### Haplotype divergence

The percentage of nucleotide sequence identity was calculated for each pair of hom(oe)ologous and orthologous genes (Table 3). Among sugarcane hom(oe)ologous alleles, sequence identity was very high, i.e. ranging from 78.8% to 100%, with an average of 95.9% (97.7% when not taking gaps into account), for the coding sequence, and from 39.3% to 100%, with an average of 87.5% (96.9% when not taking gaps into account), for the aligned part of the introns. Between sugarcane and sorghum, the average identity was high, with 91.6% for the coding sequence and 72.8% for the aligned part of introns (96.4% and 88.0% when not taking gaps into account, respectively). Between sugarcane and rice, the average identity was lower as expected, with 71% for the coding sequence and 38% for the aligned part of the intron.

Haplotype differentiation was analyzed by comparing sequences of hom(oe)ologous alleles (Table 3) and building phylogenetic trees and haplotype networks (data not shown). A synthetic representation of the results is shown in Fig. 4. The maximum divergence between two alleles within a locus ranged from 1.2 to 9.6 Myr for the segmental duplicated loci 11b and 11a, respectively. Among nonsegmental duplicated loci, this range was from 1.7 to 5.1 Myr. The most divergent alleles fell into haplotypes I, III and part of haplotype VII, and the most similar into



**Fig. 1** Comparison of the physical organization between seven hom(eo)ologous sugarcane bacterial artificial chromosome (BAC) clones, and with sorghum and rice orthologous segments in the *Brv1* region. Genes are indicated by black boxes, long terminal repeat (LTR) retrotransposons by gray boxes (insertion times are indicated inside boxes), non-LTR retrotransposons by gray striped boxes, transposons by gray dotted boxes and proteinase inhibitor and proteinase inhibitor by diamonds. Colinear genes are connected by shaded areas and colinear transposable elements (TEs) by dashed lines. Genes are numbered according to Table 2. TE family names are abbreviated as: Rh (Rhum), Ti (TIPunch), Da (Daiquiri), Ca (Caipirinha), Ch (Cachaca), Co (Colada), PI (Planteur) and Mo (Mojito).

**Table 2** List of syntenic genes on sugarcane bacterial artificial chromosome (BAC) clones

No.	Protein giving highest BLAST score	Best BLASTX E-value	Syntenic orthologous sorghum locus <sup>1</sup>
1	<i>Oryza sativa</i> conserved hypothetical protein (CHP) (Q2QZU3)	3.E-18	Sb04g001890
2	<i>Arabidopsis thaliana</i> histone lysine N-methyltransferase ATXR6 (Q9FNE9)	1.E-123	Sb04g001900
3	<i>O. sativa</i> putative arginine/serine-rich splicing factor (Q6Z725)	1.E-80	Sb04g001910
4	<i>A. thaliana</i> dimethyladenosine transferase (O65090)	7.E-89	Sb04g001913
5	<i>O. sativa</i> Ulp1 protease family, C-terminal catalytic domain containing protein (Q2QME9)	1.E-142	Sb04g001916
6	<i>O. sativa</i> cell division control protein 2 homolog 2 (P29619)	1.E-144	Sb04g001920
7	<i>A. thaliana</i> shrunken seed protein (Q85851)	5.E-88	Sb04g001930
8	<i>O. sativa</i> conserved hypothetical protein (CHP) PO575F10.13 (Q62719)	0.E + 00	Sb04g001940
9	<i>O. sativa</i> tyrosine-specific protein phosphatase-like (Q6ZHH2)	5.E-26	–
10	<i>Triticum aestivum</i> serine/threonine protein kinase pelle fragment (Q5BQ31)	8.E-15	–
11a	<i>Malus domestica</i> NADP-dependent D-sorbitol-6-phosphate dehydrogenase (P28475)	3.E-123	–
11b	<i>Malus domestica</i> NADP-dependent D-sorbitol-6-phosphate dehydrogenase (P28475)	3.E-125	Sb04g001950
12	<i>O. sativa</i> endoglucanase 4 precursor (Q6Z715)	1.E-152	Sb04g001960
13	<i>O. sativa</i> hypothetical protein (HP) PO662B01.10 (Q5VQ47)	1.E-38	–
14	<i>O. sativa</i> hypothetical protein (HP) H0821G03.11 (Q25A05)	2.E-04	–

<sup>1</sup>, Loci numbers are available at <http://www.phytozome.net/>, v1.0 release, comprising the Sbi1 assembly and Sbi1.4 gene set.

haplotypes II, IV, V and VI. The simplest interpretation is that these two sets of haplotypes (II, IV, V and VI vs I and III) could originate from distinct genomes, that is *S. officinarum* and *S. spontaneum*. The variation along haplotype VII suggests a more complex origin. The most divergent haplotypes or haplotype fragments could tentatively be attributed to *S. spontaneum* on the basis of the much broader molecular differentiation observed between *S. officinarum* and *S. spontaneum* than within *S. officinarum*, as well as the greater diversity and heterozygosity within *S. spontaneum* than within *S. officinarum* (Lu *et al.*, 1994; Jannoo *et al.*, 1999). It is noteworthy that, with the exception of haplotype VII, the presence vs absence of gene 11a and the distribution of the few colinear TEs corroborate the sequence differentiation pattern.

On the basis of the average gene synonymous substitutions between all sugarcane haplotypes and the sorghum and rice orthologous sequences, we estimated that sugarcane and sorghum diverged around 6–9 Mya and sorghum and rice around 43 Mya, which is in agreement with the previously published dates of 8–9 Mya (Jannoo *et al.*, 2007) and 40–50 Mya (Paterson *et al.*, 2004; Bowers *et al.*, 2005), respectively.

### Selective constraints on hom(oe)ologous genes

The synonymous ( $K_s$ ) and nonsynonymous ( $K_a$ ) substitution rates, as well as the  $K_a : K_s$  ratios, were calculated for exons of all pairs of hom(oe)ologous alleles (Fig. 2). For nine of the 14 genes compared (genes 2, 3, 4, 5, 6, 7, 8, 11a and 14),  $K_a : K_s$  values under unity were obtained for all pairs of hom(oe)ologous alleles. For seven of these genes, significant  $P$  values for the  $Z$ -test (below 0.05 and down to below 0.001) were observed for several pairs of hom(oe)olo-

gous alleles. In the other cases, the results obtained were not statistically significant because of the small number of compared sequences and/or a lack of substitutions. For genes 9 and 13, some  $K_a : K_s$  values were close to unity, but they were not statistically significant. These results indicated that the majority of hom(oe)ologous alleles were under purifying selection (Fig. 2). The only genes for which some pairs of hom(oe)ologous alleles showed a  $K_a : K_s$  ratio above unity with, in a few cases, a  $Z$ -test below 0.05, concerned gene fragment 10 and pairs of hom(oe)ologous alleles within segmental duplications: genes 11 (duplication) and gene 12 (triplication).

Gene 11 was duplicated on four of the hom(oe)ologous sugarcane haplotypes, whereas the other haplotypes contained only gene 11b. Sorghum and rice have only one copy of this gene and phylogenetic analyses showed that it had greater similarity to gene 11b (Fig. 3). Alleles of gene 11a displayed four- to five-fold more substitutions than alleles of the other genes in the sugarcane region. However, in spite of this excess of substitutions, a very high purifying selection pressure (with most  $P$  values below 0.001) was observed for gene 11a, whereas more relaxed selective pressures were observed for gene 11b.

### Comparison of TE content within sugarcane haplotypes and with sorghum and rice

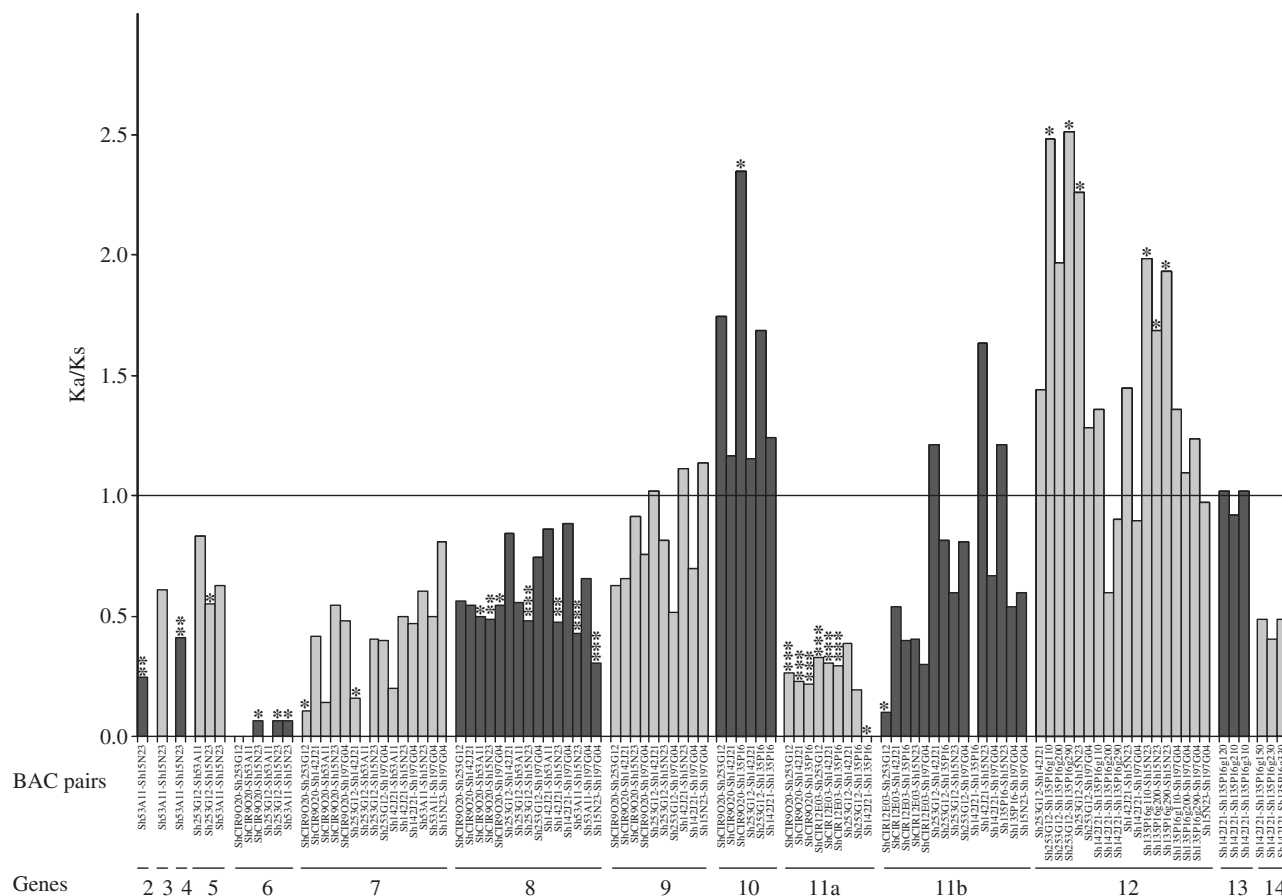
Sixty-six large TEs were annotated on BAC clones. They represented an average of 35% of all BAC sequences (between 15% and 54%). Most TEs had high similarities with previously reported families, but 21% were new. We attributed family names to the most abundant sugarcane elements, and their classification is described in Table 4. LTR retrotransposons were the most frequent elements,

**Table 3** Percentage of identity between hom(oe)ologous sugarcane genes (a) and orthologous sorghum–sugarcane genes (b)

Gene	No. of sequences	Alignment length (bp)	Exon length (bp)	Intron length (bp)	Percentage identity of exons			Percentage identity of introns		
					Average	Max value	Min value	Average	Max value	Min value
(a)										
2. Histone-lysine <i>N</i> -methyltransferase	2	1195	1101	94	97.6	97.6	97.6	100.0	100.0	100.0
3. Arginine/serine-rich splicing factor	2	3449	975	2474	96.7	96.7	96.7	89.8	89.8	89.8
4. Dimethyladenosine transferase	2	3965	1002	2963	96.8	96.8	96.8	74.4	74.4	74.4
5. Putative ulp1 protease	3	7944	2586	5358	98.4	98.5	98.3	90.6	97.7	86.6
6. Cell division control protein 2 homolog 2	4	4315	882	3433	99.5	100.0	99.2	91.6	98.7	85.2
7. Putative shrunken seed protein	6	1872	966	906	99.2	99.9	98.6	97.6	99.9	95.8
8. Conserved hypothetical protein	6	3021	2445	576	96.3	99.3	92.4	96.8	99.6	93.6
9. Putative tyrosine-specific protein phosphatase	5	1662	1662	0	90.3	99.3	84.6	–	–	–
10. Serine/threonine kinase (fragment)	4	863	411	452	88.8	93.1	84.4	69.2	100.0	40.2
11a. NADP-dependent D-sorbitol-6-phosphate dehydrogenase	5	1953	681	1272	92.8	99.7	87.4	66.5	98.2	39.3
11b. NADP-dependent D-sorbitol-6-phosphate dehydrogenase	6	5113	957	4156	98.3	99.7	95.8	76.3	95.5	59.2
12. Endoglucanase 4 precursor	7 <sup>1</sup>	2107	1757	350	90.0	99.8	78.8	85.4	99.4	71.1
13. Hypothetical protein	4 <sup>1</sup>	2791	1818	973	99.0	99.8	98.0	99.4	99.9	99.3
14. Hypothetical protein	4 <sup>1</sup>	638	552	86	99.0	100.0	98.0	100.0	100.0	100.0
Average	4	2921	1271	1650	95.9	98.6	93.3	87.5	96.4	79.6
(b)										
2. Histone-lysine <i>N</i> -methyltransferase	3	1195	1101	94	94.1	94.2	94.1	87.2	87.2	87.2
3. Arginine/serine-rich splicing factor	3	3590	975	2615	94.4	95.6	93.5	84.9	87.7	82.1
4. Dimethyladenosine transferase	3	4019	1032	2987	90.5	92.3	89.3	62.8	64.5	61.2
5. Putative ulp1 protease	4	8058	2607	5451	89.7	89.8	89.7	80.5	83.5	74.5
6. Cell division control protein 2 homolog 2	5	4508	891	3617	95.2	95.4	95.1	74.3	75.7	70.7
7. Putative shrunken seed protein	7	3696	967	2729	94.1	94.6	94.0	83.9	86.5	78.3
8. Conserved hypothetical protein	7	3137	2445	692	92.7	93.4	91.1	87.6	88.7	85.4
9. Putative tyrosine-specific protein phosphatase	6	210	210	0	82.4	82.9	81.9	–	–	–
11b. NADP-dependent D-sorbitol-6-phosphate dehydrogenase	7	5296	966	4330	94.0	94.9	92.2	52.2	58.5	47.5
12. Endoglucanase 4 precursor	8 <sup>1</sup>	1852	1644	208	88.5	91.8	80.8	41.5	49.4	39.4
Average	5	3556	1284	2272	91.6	92.5	90.2	72.8	75.7	69.6

<sup>1</sup>, With three paralogous genes in the Sh135P16 bacterial artificial chromosome.





**Fig. 2**  $K_a : K_s$  values for each gene from bacterial artificial chromosome (BAC) pair comparisons. Pairwise comparisons with a Z-test gave  $P$  values of \*,  $P \leq 0.05$ ; \*\*,  $P \leq 0.01$ ; \*\*\*,  $P \leq 0.001$ .

representing 65% of all TEs. They all belonged to two superfamilies, namely Ty3-Gypsy (58%) and Ty1-Copia (42%). Non-LTR retrotransposons represented 17% of all TEs; they all belonged to the LINE superfamily. DNA transposons (TE class II) represented 18% of all elements, with two-thirds belonging to the CACTA superfamily.

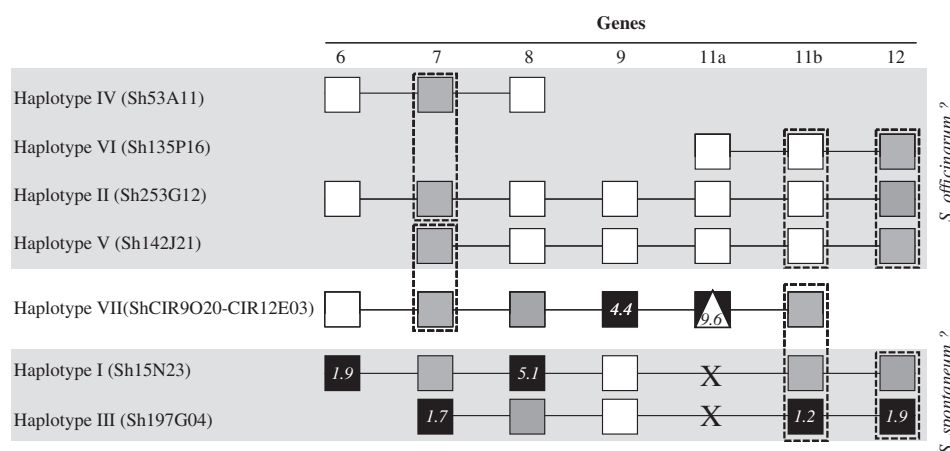
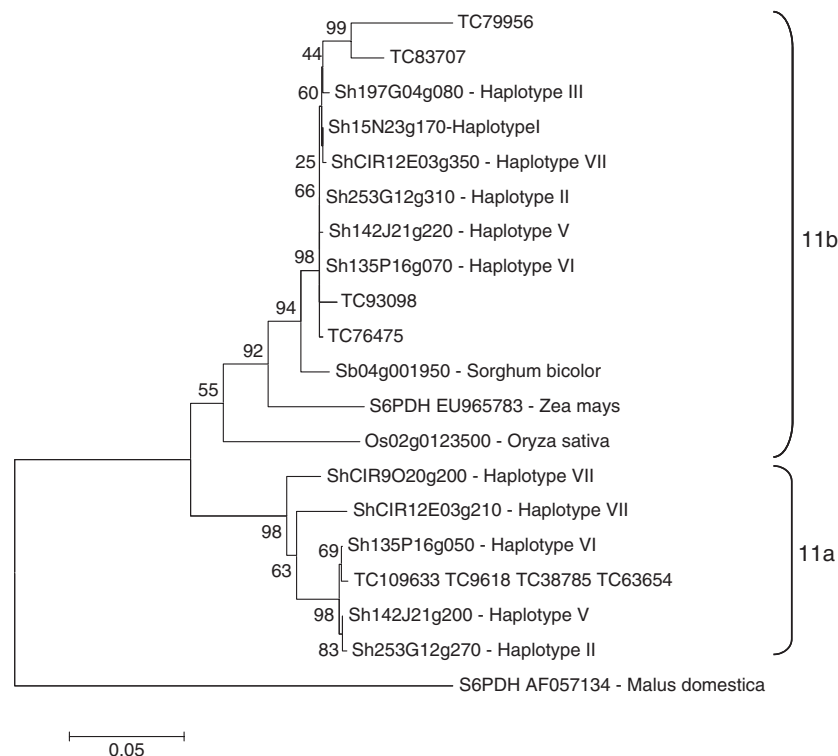
Twenty percent of the TEs were complete, including 12 LTR retrotransposons and one transposon. Insertion times were calculated for all complete LTR retrotransposons and ranged from 0 to 1.58 Mya (Table 4). The vast majority of TEs were located in intergenic regions, except for eight that were inserted in introns of one allele of gene 11a and four alleles of gene 11b. No large TE was observed between genes 1 and 12 in the sorghum and rice orthologous regions. Most TEs were found in several hom(o)ologous haplotypes, and no evident specific distribution of TEs between haplotypes was noted. They displayed hardly any colinearity, with their position not being conserved across haplotypes, except in three cases. The exceptions concerned a LTR retrotransposon of the Rhum family, which was present on haplotype V (two elements in tandem with one shared LTR), haplotype II (fragment) and haplotype VII

(fragment) (Fig. 1). The second case concerned a LTR retrotransposon of the Colada family, which was complete on haplotype II and present as a fragment on haplotype IV. The third case concerned a fragment of a transposon found at the same position on haplotypes V and VI.

## Discussion

Our study was based on BAC clones from a modern sugarcane cultivar. Typically, modern cultivars have a genome made up of 75–85% of chromosomes from the autopolyploid species *S. officinarum* and 15–25% of chromosomes from the autopolyploid species *S. spontaneum* (D'Hont *et al.*, 1996; Piperidis *et al.*, 2010). These two types of chromosome were juxtaposed very recently (a century ago) through breeding, after their evolution within the autopolyploid parental species. Dating the origins of polyploid events is difficult (Doyle & Egan, 2009), particularly for sugarcane: autopolyploidy implies a lack of initial differentiation (as opposed to allopolyploidy); it is likely to be gradual in such a high polyploid and its diploid progenitors have become extinct. On the basis of the molecular clock of

**Fig. 3** Phylogenetic relationships of genes 11 encoding sorbitol-6-phosphate dehydrogenase (S6PDH) proteins. Gene 11 coding sequences from sugarcane bacterial artificial chromosomes (BACs), tentative consensus (TC) available in the TIGR sugarcane database, sorghum and rice were used to construct the tree by the neighbor-joining method implemented in MEGA4 software. The robustness of the tree topology was assessed with 1000 bootstrap replicates. The scale bar represents the relative genetic distance (number of substitutions per nucleotide). The coding sequences of *S6PDH* from *Zea mays* and *Malus domestica* (GenBank numbers EU965783 and AF057134, respectively) were used to root the tree. *S6PDH* genes were subdivided into at least two groups, namely 11a and 11b. ShCIR9O20g200 corresponds to a fragment of a second copy of gene 11a present on haplotype VII at the extremity of BAC ShCIR9O20, but interrupted by the cloning site of BAC.



**Fig. 4** Schematic representation of differentiation among sugarcane haplotypes based on hom(oe)ologous allele sequence comparison (see text). Each allele is represented by a square. For each locus, the most divergent allele is marked in black and its theoretical divergence time (highest estimate observed in Myr) is indicated in italics. All alleles that fall into groups (of at least three) with all values lower than one-third of this maximum divergence time are marked by white squares. When the phylogenetic trees were not degenerate, the alleles of the same branch (relating to the same internal node) were placed in vertical dotted boxes, providing clues for classification (loci 7, 11b and 12). The white triangle in the black square for locus 11a indicates an insertion. The 'x' mark indicates absence of the gene.

*Adh* gene sequences, we estimated that the closest sugarcane diploid relative identified so far, that is *Narenga porphyrocoma* (Al-Janabi *et al.*, 1994), diverged from sugarcane at 2.5 Mya (A. D'Hont, unpublished), and that *S. spontaneum* and *S. officinarum* diverged at 1.5–2 Mya (Jannoo *et al.*, 2007). Polyploidy probably occurred after this divergence as these species have different basic chromo-

somes numbers (D'Hont *et al.*, 1998). Both *S. spontaneum* and the wild progenitor of *S. officinarum* (*S. robustum*) feature an impressive series of high polyploids. *Saccharum* species are therefore considered to be relatively ancient polyploids. Between the evolution as autopolyploid natural species and the short history as interspecific breeding materials, limited to the last century and a few meioses

**Table 4** List and classification of large transposable elements (TEs) on sugarcane bacterial artificial chromosome (BAC) clones

Class	Order	Superfamily	Classification <sup>1</sup> _ Sugarcane family	Locus_tag	Structure	TSD	T(Mya) <sup>2</sup>	SE <sup>3</sup>	RepBase highest similarity
LTR-Retrotransposon	LTR	Copia	RLC_Rhum	Sh135P16_te010	solo_LTR				
				Sh142J21_te040 *	complete	GTCG	0.21	0.16	
				Sh142J21_te050 *	complete	GTCG	0.19	0.15	
				Sh253G12_te010	fragment				Copia82 & 37-ZM_LTR
				Sh253G12_te015	fragment				
				Sh253G12_te040 *	fragment				
			RLC_Caipirinha	ShCIR012E03_te100 *	fragment				
				Sh053A11_te010	complete	ACGGT	0.00	0.00	
				Sh053A11_te030	fragment				
				Sh053A11_te035	fragment				Copia22-ZM_LTR
				Sh253G12_te027	fragment				
				ShCIR012E03_te075	fragment				
				ShCIR012E03_te110	fragment				
			RLG_TiPunch	ShCIR009O20_te050	fragment				Copia-10_SB_LTR
				ShCIR012E03_te040	fragment				no significant hit
				Sh197G04_te020	complete	GGGGC	1.35	0.35	Copia29-ZM_LTR
				Sh015N23_te040	complete	CCGTG	0.08	0.08	no significant hit
				Sh135P16_te030	fragment				Copia-76_SB_LTR
				Sh135P16_te080	fragment				
		Gypsy	RLG_Colada	Sh194G04_te025	fragment				
				Sh253G12_te025	fragment				
				ShCIR009O20_te010	complete	TCACC	0.77	0.27	Gypsy-126_SB_LTR
				ShCIR009O20_te015	fragment				
				ShCIR009O20_te035	fragment				
				ShCIR009O20_te030	fragment				
			RLG_Cachaca	ShCIR012E03_te095	fragment				
				Sh053A11_te020 *	fragment				
				Sh142J21_te030	fragment				
				Sh197G04_te030	fragment				
				Sh197G04_te040	fragment				
				Sh253G12_te020 *	complete	GTCAT	1.58	0.38	Gypsy-22 & 131_SB_LTR
Non-LTR-Retrotransposon	LINE	L1	RLG_Planteur	Sh253G12_te026	fragment				
				ShCIR012E03_te096	fragment				
				ShCIR012E03_te070	complete	ATATT	1.27	0.35	
				Sh142J21_te010	complete	-	0.38	0.19	
				Sh197G04_te050	complete	GCCAG	0.46	0.19	
				Sh197G04_te060	fragment				Gypsy-29_SB_LTR
			RLG_Daiquiri	ShCIR009O20_te040	fragment				
				ShCIR009O20_te045	fragment				
				Sh015N23_te010	complete	CAACG	0.77	0.27	
				Sh253G12_te028	fragment				no significant hit
				ShCIR012E03_te020	fragment				
				Sh142J21_te020	complete	ATCGG	0.04	0.04	Gypsy-14_SB_LTR
			RIL_NB OS	Sh015N23_te080	fragment				
				Sh197G04_te010	fragment				
				ShCIR012E03_te060	fragment				RTE1_ZM
				ShCIR012E03_te090	fragment				
				Sh015N23_te050	fragment				LINE1-22_ZM
				Sh015N23_te060	fragment				no significant hit
DNA-transposon	TIR	CACTA	DTC_Mojito	Sh197G04_te080	fragment				LINE1-7_SB
				ShCIR012E03_te010	fragment				LINE1-40_ZM
				ShCIR012E03_te030	fragment				no significant hit
				Sh197G04_te070	fragment				no significant hit
				Sh142J21_te060	fragment				
				Sh135P16_te040	fragment				
			hAT	Sh135P16_te050	fragment				no significant hit
				Sh135P16_te060	fragment				
				Sh135P16_te020	fragment				EnSpm-20_SB
				Sh135P16_te070	fragment				EnSpm-14_SB
				Sh135P16_te100	fragment				no significant hit
				Sh253G12_te060	fragment				ZhAT14_ZM
Helitron	helitron			ShCIR012E03_te050	fragment				
				Sh135P16_te090	fragment				Helitron-N4_SB
				Sh253G12_te030	fragment				
				ShCIR012E03_te080	complete				Helitron-2_SB

LTR, long terminal repeat. TIR, terminal inverted repeat.

<sup>1</sup>, Classification system according to Wicker *et al.* (2007). We have attributed a family name at the most frequent transposable TEs.

<sup>2</sup>, Insertion time in million years.

<sup>3</sup>, Standard error.

\*, Syntenic location (see Fig. 1).

( $\leq 7$ ), it is likely that differentiation among hom(oe)ologous haplotypes extracted from a modern sugarcane (here R570) essentially reflects the evolutionary dynamics within the two autopolyploids.

In Jannoo *et al.* (2007), we compared two homoeologous haplotypes (BAC clones bearing *Adh1* genes) from sugarcane

cultivar R570, one originating from *S. officinarum* and one from *S. spontaneum*. In the present study, we compared seven haplotypes from a second region of the same cultivar. The species' origin of the individual haplotypes was not determined precisely, but could tentatively be inferred (Fig. 4), probably featuring the two origins from *S. officinarum* and



*S. spontaneum*. Hence, the present comparisons involved homologous and homoeologous haplotypes.

### High gene conservation in the autopolyploid genome context of sugarcane

We observed a remarkably high general colinearity of genes between sugarcane haplotypes and high gene structure and sequence conservation of homologous and homoeologous alleles. The general colinearity is disturbed by only a few segmental duplications. Strikingly, all the hom(oe)ologous genes (except gene fragment 10) were predicted, based on their structure, to be functional, although gene expression still needs to be verified experimentally. This was noted by Jannoo *et al.* (2007) between two homoeologous haplotypes, and is extended here for the first time to homologous haplotypes. This high retention of redundant functional genes contrasts with the situation in paleopolyploids (Blanc & Wolfe, 2004; Thomas *et al.*, 2006; Freeling, 2009; Throude *et al.*, 2009) and in some more recent allopolyploid species, such as wheat (Ozkan *et al.*, 2001; Chantret *et al.*, 2005) and *Tragopogon* (Tate *et al.*, 2006; Buggs *et al.*, 2010), where important gene eliminations and pseudogenizations have been observed. However, it is in line with the situation in some other allopolyploids, such as *Brassica* and *Gossypium*, for which a high level of gene conservation and colinearity have also been reported (Grover *et al.*, 2004; Cheung *et al.*, 2009). When duplicated genes are conserved, they can be under relaxed selection, as exemplified by the case of the *MONOCULM1* region in *Oryza minuta*, an allotetraploid rice (BBCC) estimated to be of recent origin (0.4 Mya) (Lu *et al.*, 2009). Interestingly, our analysis in the sugarcane *Bru1* region (this study) and *Adh1* regions (Supporting information Table S1) suggests, conversely, that purifying selection acts on most genes. The only exceptions concerned genes comprising segmental duplications.

This high hom(oe)allele conservation, despite the extreme gene redundancy observed in sugarcane, raises the question of the mechanisms involved. The mode of chromosome pairing may have a marked impact on genome evolution. In particular, the recurrent random assortment of chromosomes at meiosis in autopolyploids may counterselect departure from the functional stability of all alleles, because it may give rise to individuals (and gametes) lacking a functionally complete gene set. Le Comber *et al.* (2010), using computer simulation, suggested that subfunctionalization speeds up the transition from polysomic to disomic inheritance and also acts to maintain genes in syntenic blocks. Analysis of the expression of the hom(oe)alleles identified in this study should be undertaken to test for evidence of subfunctionalization and to further investigate this hypothesis. Genetic mapping studies have shown that chromosome assortment results from general polysomy in *S. spontaneum* (Al-Janabi *et al.*, 1993) and from polysomy

with some cases of preferential pairing in *S. officinarum* (Aitken *et al.*, 2007) and in modern sugarcane cultivars (Grivet *et al.*, 1996; Hoarau *et al.*, 2001; Jannoo *et al.*, 2004). These observations may be signs of initiation of transition from polysomic to disomic inheritance.

One of the genes duplicated through segmental duplication, gene 11, corresponds to S6PDH, which is a key enzyme in sugar metabolism for sorbitol biosynthesis. The duplication apparently occurred after the divergence between sugarcane and sorghum lineages, and possibly during the differentiation of *Saccharum* species, as it is absent from some of the haplotypes and confined to those tentatively attributed to *S. officinarum* (the sugar-producing species). The two copies have diverged; gene 11a has accumulated an important number of substitutions, but has evolved under very high purifying selection pressure, whereas more relaxed selective pressures were observed for gene 11b. These results suggest that segmental duplication has allowed the diversification of gene 11a, possibly towards an important specific function in sugarcane.

The contrasting result observed with regard to the evolution of duplicated genes suggests that, in our model, 'horizontal' duplications (segmental duplications), but not 'vertical' duplications (whole genome duplications through polyploidization), allow for diversification.

### High TE variation between sugarcane haplotypes

The most represented TE families were Gypsy 'TiPunch' (12%) and Copia 'Rhum' (10%). These two retrotransposon families were also found to be the most abundant TEs in the sugarcane genome (*S. officinarum*) when analyzing about 1.45 Mb of random Sanger genomic sequence data (J. De Barry, J. L. Bennetzen, pers. comm., University of Georgia, USA). A survey of TEs in 260 781 sugarcane ESTs (Vettore *et al.*, 2003) revealed a different distribution, with 53% transposons, 46.4% LTR retrotransposons and the absence of non-LTR retrotransposons. The most abundant retrotransposon superfamily in these ESTs was Copia (30%), with the Hopscotch family being the most represented element (Rossi *et al.*, 2001; De Araujo *et al.*, 2005). This suggested a nonrandom TE distribution in the expressed part of the genome.

In contrast with gene conservation and colinearity, we observed a general absence of colinearity for TEs among sugarcane haplotypes. This absence of colinearity, also noted by Jannoo *et al.* (2007) between two homoeologous haplotypes in the *Adh1* region, was confirmed here in a second region extended to homologous haplotypes. In addition, the sugarcane sequence appears to be expanded compared with sorghum and rice orthologous regions because of the accumulation of TEs between genes. Indeed, rice and sorghum displayed no large TEs between genes 1 and 12 in the present study and between genes 1 and 14 in the *Adh1* region.

The TE fraction of the genome in plants is known to be variable in size and very dynamic because of transposition and unequal recombination (homologous and illegitimate) (Ma *et al.*, 2004; Vitte & Panaud, 2005). In some allopolyploids, major structural changes in the TE genome fraction have been observed after allopolyploidization, including TE proliferation and TE loss through recombination (review by Parisod *et al.*, 2010). When they occur, these changes are generally attributed to interspecific hybridization (genomic shock) rather than polyploidy *per se* (McClintock, 1984). In our case, the interspecific configuration that was established by modern breeding is very recent (a century ago), and the insertion times of LTR retrotransposons that could be estimated were dated much earlier. Therefore, the diversity of TE patterns is unlikely to have resulted mainly from the recent interspecific hybridization.

We can speculate that polysomy and the high effective population size associated with autopolyploidy may impact on TE dynamics, and that the presence of TEs close to genes may facilitate a diversification of hom(oe)allele expression patterns. These questions warrant further investigation.

## Conclusion

This is the first analysis of the fine structural organization of a set of homologous and homoeologous haplotypes in sugarcane, and is the first of its kind in a highly polyploid genome. Despite an extreme level of gene redundancy, our analysis revealed strikingly high retention of redundant functional genes. In contrast, a general absence of TE colinearity among hom(oe)ologous haplotypes was observed, illustrating a dynamic expansion of TEs in the founder autopolyploid *Saccharum* species compared with sorghum. These results reinforce the general trend emerging from recent studies, indicating the diverse (depending on lineages) and nuanced effect of polyploidy on genome dynamics (Doyle *et al.*, 2008; Parisod *et al.*, 2010).

Our results also provide guidelines for future sequencing and assembling strategies that are presently being discussed within the Sugarcane Genome Sequencing Initiative (SUGESI, <http://sugarcanegenome.org/>). In particular, our results reveal high gene colinearity among hom(oe)ologous haplotypes, suggesting that one haplotype can serve as reference for the other hom(oe)ologous haplotypes with regard to gene content. They also confirm the high gene microlinearity between sorghum and sugarcane (Jannoo *et al.*, 2007; Wang *et al.*, 2010), thus making the sorghum sequence a good template to facilitate the assembly of the gene-rich part of the sugarcane genome. This study also has practical implications for the development of new molecular marker technologies. For example, to optimize the development of haplotype-specific markers, it could be more efficient to favor technologies based on TE variation, such as insertion

site-based polymorphism markers (Paux *et al.*, 2010), rather than those based on gene variation.

Sugarcane has been recognized as one of the world's most efficient crops in solar energy conversion and as having the most favorable input : output ratios. Our data suggest the presence of broad sets of functional homologous alleles in its genome, which could explain its unique efficiency, particularly its high phenotypic plasticity and wide adaptation.

## Acknowledgements

We thank Douglas D. Silva, Magdalena Rossi, Erika M. de Jesus and Nilo Saccaro-Junior from the GaTE Laboratory, University of Sao Paulo for their help in annotating TEs, and Nabila Yahiaoui and three anonymous referees whose contributions improved the manuscript. This work was funded by Genoscope (AAP2005) and Cirad.

## References

- Aitken KS, Jackson PA, McIntyre CL. 2007. Construction of a genetic linkage map for *Saccharum officinarum* incorporating both simplex and duplex markers to increase genome coverage. *Genome* 50: 742–756.
- Albertin W, Brabant P, Catrice O, Eber F, Jenczewski E, Chevre AM, Thielllement H. 2005. Autopolyploidy in cabbage (*Brassica oleracea* L.) does not alter significantly the proteomes of green tissues. *Proteomics* 5: 2131–2139.
- Al-Janabi SM, Honeycutt RJ, McClelland M, Sobral BW. 1993. A genetic linkage map of *Saccharum spontaneum* L. 'SES 208'. *Genetics* 134: 1249–1260.
- Al-Janabi SM, McClelland M, Petersen C, Sobral BW. 1994. Phylogenetic analysis of organellar DNA sequences in the Andropogoneae: Saccharinae. *Theoretical and Applied Genetics* 88: 933–944.
- Blanc G, Wolfe KH. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16: 1667–1678.
- Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC *et al.* 2005. Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. *Proceedings of the National Academy of Sciences, USA* 102: 13206–13211.
- Brandes E. 1956. Origin, dispersal and use in breeding of the Melanesian garden sugarcane and their derivatives, *Saccharum officinarum* L. *Proceedings of the International Society of Sugar Cane Technologists* 9: 709–750.
- Bremer G. 1922. Een cytologisch onderzoek van eenige soorten en soortsbastarden van het geslacht *Saccharum*. *Arch Suikerindust Nederl Indie Meded Proefsta Java Suikerindust* 1–112. (English translation in *Genetica* 5: 97–148; 273–326.
- Buggs RJ, Chamala S, Wu W, Gao L, May GD, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB. 2010. Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Molecular Ecology* 19 (Suppl. 1): 132–146.
- Burner DM, Legendre BL. 1994. Cytogenetic and fertility characteristics of elite sugarcane clones. *Sugarcane* 1: 6–10.
- Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P *et al.* 2005. Molecular basis of

- evolutionary events that shaped the *hardness* locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell* 17: 1033–1045.
- Cheung F, Trick M, Drou N, Lim YP, Park JY, Kwon SJ, Kim JA, Scott R, Pires JC, Paterson AH *et al.* 2009. Comparative analysis between homoeologous genome segments of *Brassica napus* and its progenitor species reveals extensive sequence-level divergence. *Plant Cell* 21: 1912–1928.
- Church SA, Spaulding EJ. 2009. Gene expression in a wild autopolyploid sunflower series. *Journal of Heredity* 100: 491–495.
- Consortium U. 2009. The universal protein resource (UniProt) 2009. *Nucleic Acids Research* 37: D169–D174.
- Cuadrado A, Acevedo R, Moreno Diaz de la Espina S, Jouve N, De la Torre C. 2004. Genome remodelling in three modern *S. officinarum* × *S. spontaneum* sugarcane cultivars. *Journal of Experimental Botany* 55: 847–854.
- De Araujo PG, Rossi M, de Jesus EM, Saccaro NL Jr, Kajihara D, Massa R, de Felix JM, Drummond RD, Falco MC, Chabregas SM *et al.* 2005. Transcriptionally active transposable elements in recent hybrid sugarcane. *Plant Journal* 44: 707–717.
- De Silva HN, Hall AJ, Rikkerink E, McNeilage MA, Fraser LG. 2005. Estimation of allele frequencies in polyploids under certain patterns of inheritance. *Heredity* 95: 327–334.
- D'Hont A. 2005. Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenetics and Genome Research* 109: 27–33.
- D'Hont A, Glaszmann JC. 2001. Sugarcane genome analysis with molecular markers, a first decade of research. *Proceedings of the International Society of Sugarcane Technology* 24: 556–559.
- D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann JC. 1996. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Molecular and General Genetics* 250: 405–413.
- D'Hont A, Ison D, Alix K, Roux C, Glaszmann JC. 1998. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome* 41: 221–225.
- Doyle JJ, Egan AN. 2009. Dating the origins of polyploidy events. *New Phytologist* 186: 73–85.
- Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* 42: 443–461.
- Dufour P, Deu M, Grivet L, D'Hont A, Paulet F, Bouet A, Lanaud C, Glaszmann JC, Hamon P. 1997. Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. *Theoretical and Applied Genetics* 94: 409–418.
- Fawcett JA, Maere S, Van de Peer Y. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proceedings of the National Academy of Sciences, USA* 106: 5737–5742.
- Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology* 60: 433–453.
- Gaut BS, Morton BR, McCaig BC, Clegg MT. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcl*. *Proceedings of the National Academy of Sciences, USA* 93: 10274–10279.
- Grivet L, D'Hont A, Dufour P, Hamon P, Roques D, Glaszmann JC. 1994. Comparative genome mapping of sugar cane with other species within the Andropogoneae tribe. *Heredity* 73: 500–508.
- Grivet L, D'Hont A, Roques D, Feldmann P, Lanaud C, Glaszmann JC. 1996. RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid. *Genetics* 142: 987–1000.
- Grivet L, Glaszmann JC, D'Hont A. 2006. Molecular evidence of sugarcane evolution and domestication. In: Motley TJ, ed. *Darwin's harvest: new approaches to the origins, evolution and conservation of crops*. New York, NY, USA: Columbia University Press, 49–66.
- Grover CE, Kim H, Wing RA, Paterson AH, Wendel JF. 2004. Incongruent patterns of local and global genome size evolution in cotton. *Genome Research* 14: 1474–1482.
- Guimaraes CT, Sills GR, Sobral BW. 1997. Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. *Proceedings of the National Academy of Sciences, USA* 94: 14261–14266.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.
- Hoarau JY, Offmann B, D'Hont A, Risterucci AM, Roques D, Glaszmann JC, Grivet L. 2001. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). I. Genome mapping with AFLP markers. *Theoretical and Applied Genetics* 103: 84–87.
- Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L *et al.* 2009. InterPro: the integrative protein signature database. *Nucleic Acids Research* 37: D211–D215.
- Ilic K, SanMiguel PJ, Bennetzen JL. 2003. A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. *Proceedings of the National Academy of Sciences, USA* 100: 12265–12270.
- Jackson S, Chen ZJ. 2009. Genomic and expression plasticity of polyploidy. *Current Opinion in Plant Biology* 13: 1–7.
- Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C *et al.* 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467.
- Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A. 2007. Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. *Plant Journal* 50: 574–585.
- Jannoo N, Grivet L, David J, D'Hont A, Glaszmann JC. 2004. Differential chromosome pairing affinities at meiosis in polyploid sugarcane revealed by molecular markers. *Heredity* 93: 460–467.
- Jannoo N, Grivet L, Seguin M, Paulet F, Domaingue R, Rao PS, Dookun A, D'Hont A, Glaszmann JC. 1999. Molecular investigation of the genetic base of sugarcane cultivars. *Theoretical and Applied Genetics* 99: 171–184.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenetics and Genome Research* 110: 462–467.
- Kumar S, Tamura K, Nei M. 1994. MEGA: molecular evolutionary genetics analysis software for microcomputers. *Computer Applications in the Biosciences* 10: 189–191.
- Le Comber SC, Ainouche ML, Kovarik A, Leitch AR. 2010. Making a functional diploid: from polysomic to disomic inheritance. *New Phytologist* 186: 113–122.
- Le Cunff L, Garsmeur O, Raboin LM, Pauquet J, Telismart H, Selvi A, Grivet L, Philippe R, Begum D, Deu M *et al.* 2008. Diploid/polyploid syntenic shuttle mapping and haplotype-specific chromosome walking toward a rust resistance gene (*Bru1*) in highly polyploid sugarcane ( $2n \sim 12x \sim 115$ ). *Genetics* 180: 649–660.
- Lu F, Ammiraju JS, Sanyal A, Zhang S, Song R, Chen J, Li G, Sui Y, Song X, Cheng Z *et al.* 2009. Comparative sequence analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. *Proceedings of the National Academy of Sciences, USA* 106: 2071–2076.
- Lu YH, D'Hont A, Walker DIT, Rao PS, Feldmann P, Glaszmann JC. 1994. Relationships among ancestral species of sugarcane revealed with RFLP using single copy maize nuclear probes. *Euphytica* 78: 217–226.
- Lu B, Pan X, Zhang L, Huang B, Sun L, Li B, Yi B, Zheng S, Yu X, Ding R *et al.* 2006. A genome-wide comparison of genes responsive to



- autopolyploidy in *Isatis indigotica* using *Arabidopsis thaliana* Affymetrix genechips. *Plant Molecular Biology Reporter* 24: 197–204.
- Ma J, Bennetzen JL. 2004. Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences, USA* 101: 12404–12410.
- Ma J, Devos KM, Bennetzen JL. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Research* 14: 860–869.
- Martelotto LG, Ortiz JPA, Stein J, Espinoza F, Quarín CL, Pessino SC. 2005. A comprehensive analysis of gene expression alterations in a newly synthesized *Paspalum notatum* autotetraploidy. *Plant Science* 169: 211–220.
- Mather K. 1936. Segregation and linkage in autotetraploids. *Journal of Genetics* 30: 53–78.
- McCarthy EM, McDonald JF. 2003. LTR\_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* 19: 362–367.
- McClintock B. 1984. The significance of responses of the genome to challenge. *Science* 226: 792–801.
- Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH *et al.* 1998. Detailed alignment of sorghum and rice chromosomes: comparative organization of closely related diploid and polyploid genomes. *Genetics* 150: 1663–1682.
- Nei M, Kumar S. 2000. *Molecular Evolution and Phylogenetics*. Oxford, UK/New York, NY, USA: Oxford University Press.
- Ozkan H, Levy AA, Feldman M. 2001. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* 13: 1735–1747.
- Panje R, Babu C. 1960. Studies in *Saccharum spontaneum*. Distribution and geographical association of chromosome numbers. *Cytologia* 25: 152–172.
- Parisod C, Holderegger R, Brochmann C. 2010. Evolutionary consequences of autopolyploidy. *New Phytologist* 186: 5–17.
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A *et al.* 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457: 551–556.
- Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences, USA* 101: 9903–9908.
- Paux E, Faure S, Choulet F, Roger D, Gauthier V, Martinant JP, Sourdille P, Balfourier F, Le Paslier MC, Chauveau A *et al.* 2010. Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. *Plant Biotechnology Journal* 8: 196–210.
- Piperidis G, Piperidis N, D'Hont A. 2010. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Molecular Genetics and Genomics* 284: 65–73.
- Price S. 1963. Cytogenetics of modern sugarcane. *Economic Botany* 17: 97–105.
- Rossi M, Araujo PG, Van Sluys MA. 2001. Survey of transposable elements in sugarcane expressed sequence tags (ESTs). *Genetics and Molecular Biology* 24: 147–154.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16: 944–945.
- Salse J, Abrouk M, Bolot S, Guilhot N, Courcelle E, Faraut T, Waugh R, Close TJ, Messing J, Feuillet C. 2009. Reconstruction of monocotyledonous proto-chromosomes reveals faster evolution in plants than in animals. *Proceedings of the National Academy of Sciences, USA* 106: 14908–14913.
- Schiex T, Moisan A, Rouzé P. 2001. EuGène: an eukaryotic gene finder that combines several sources of evidence. In: Gascuel O and Sagot MF, eds. *Lecture Notes in Computer Science*, Vol. 266. Germany: Springer-Verlag, 111–125.
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J *et al.* 2010. Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178–183.
- Sreenivasan TV, Ahloowalia BS, Heinz DJ. 1987. Cytogenetics. In: Heinz DJ, ed. *Sugarcane improvement through breeding*. New York, NY, USA: Elsevier Press, 211–253.
- Stupar RM, Bhaskar PB, Yandell BS, Rensink WA, Hart AL, Ouyang S, Veilleux RE, Busse JS, Erhardt RJ, Buell CR *et al.* 2007. Phenotypic and transcriptomic changes associated with potato autopolyploidization. *Genetics* 176: 2055–2067.
- Tate JA, Ni Z, Scheen AC, Koh J, Gilbert CA, Lefkowitz D, Chen ZJ, Soltis PS, Soltis DE. 2006. Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* 173: 1599–1611.
- Thomas BC, Pedersen B, Freeling M. 2006. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Research* 16: 934–946.
- Throude M, Bolot S, Bosio M, Pont C, Sarda X, Quraishi UM, Bourgis F, Lessard P, Rogowsky P, Ghesquiere A *et al.* 2009. Structure and expression analysis of rice paleo duplications. *Nucleic Acids Research* 37: 1248–1259.
- Tomkins JP, Yu Y, Miller-Smith H, Frisch DA, Woo SS, Wing R. 1999. A bacterial artificial chromosome library for sugarcane. *Theoretical and Applied Genetics* 99: 419–424.
- Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MI, Henrique-Silva F, Gigliotti EA, Lemos MV, Coutinho LL *et al.* 2003. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Research* 13: 2725–2735.
- Vitte C, Panaud O. 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetics and Genome Research* 110: 91–107.
- Wang J, Roe B, Macmill S, Yu Q, Murray J, Tang H, Chen C, Najjar F, Wiley G, Bowers J *et al.* 2010. Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. *BMC Genomics* 11: 261.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capi P, Chalouh B, Flavell A, Leroy P, Morgante M, Panaud O *et al.* 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8: 973–982.

## Supporting Information

Additional supporting information may be found in the online version of this article.

**Table S1** Selective pressures on genes belonging to the *Adh1* region (Jannoo *et al.*, 2007)

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

---

---

## **Annexe 2. Amplifications PCR et analyses microsatellites sur la collection de clones d'ADN de *Saccharum***

---

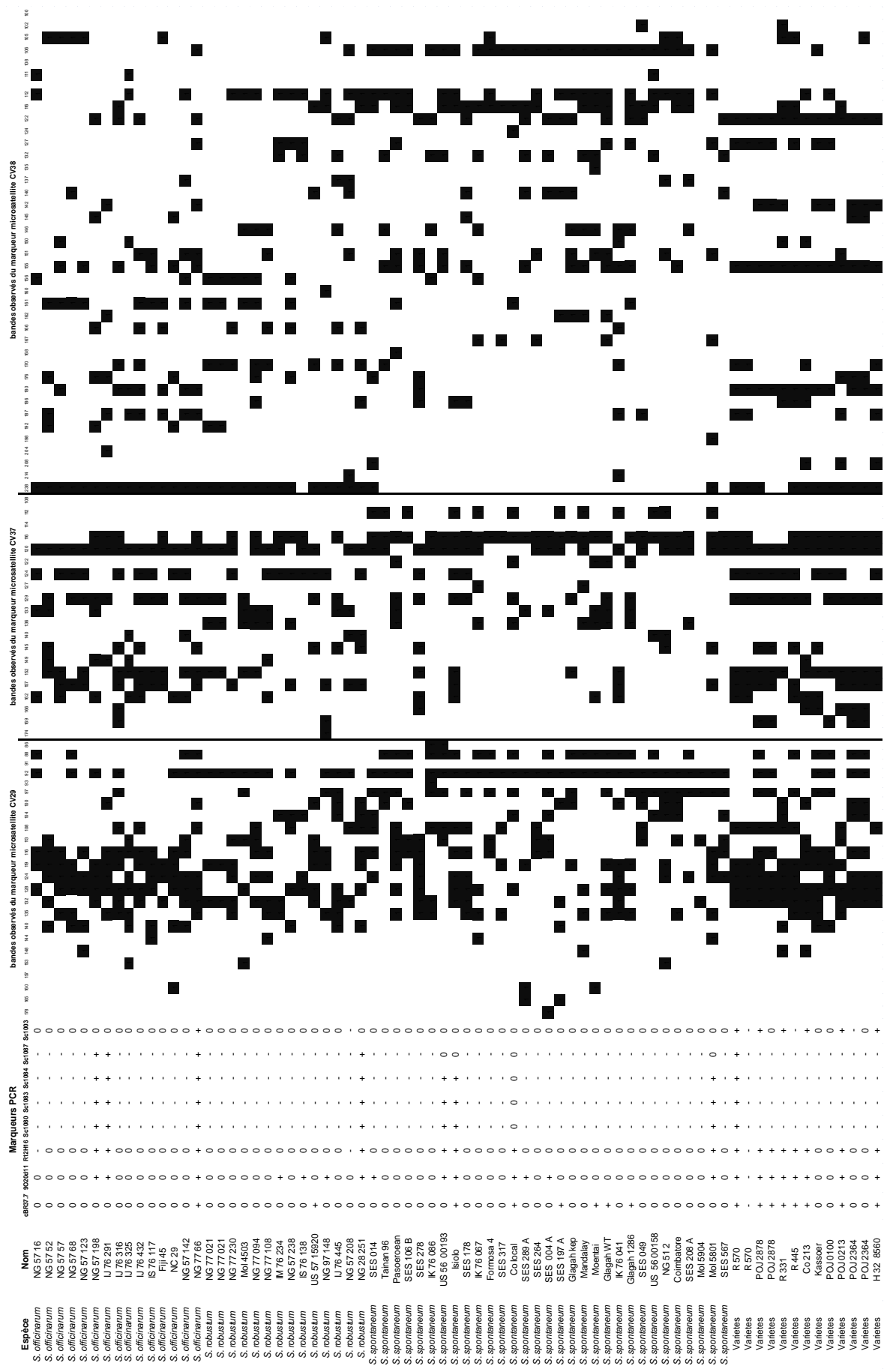
---















Les cultivars modernes de canne à sucre présentent un génome particulièrement complexe, hautement polyploïde, aneuploïde et sont issus de croisements interspécifiques entre deux espèces polyploïdes, une espèce sucrée domestiquée *Saccharum officinarum* et une espèce sauvage *Saccharum spontaneum*. Le gène majeur de résistance durable à la rouille brune, *Bru1*, a été identifié chez le cultivar de canne à sucre R570. Une approche de clonage positionnel de ce gène a été entreprise et a permis de construire une première carte physique. Cette carte comprend sept haplotypes hom(é)ologues dont un correspond à l'haplotype cible porteur du gène *Bru1*, lui-même composé de sept clones BAC qui ne se chevauchent que partiellement, laissant deux espaces non couverts. Cette situation résulte de la présence d'une insertion dans l'haplotype porteur de *Bru1*. Dans l'objectif de compléter la carte physique de l'haplotype, deux stratégies exploitant l'annotation des clones BAC ont été employées : (i) la première repose sur la conservation des gènes mise en évidence entre les différents haplotypes hom(é)ologues de la région de *Bru1* et (ii) la seconde est basée sur une marche chromosomique classique en utilisant des marqueurs flanquant les deux espaces. Ces stratégies nous ont permis de combler un des deux espaces, de couvrir partiellement le deuxième espace et de montrer que le gène de résistance se situerait dans l'insertion. Un gène candidat correspondant à une Sérine/Thréonine kinase a été identifié dans l'insertion. Des tests d'expression sous conditions d'infection seront réalisés pour valider, ou non, ce gène candidat. Parallèlement, la recherche de l'origine du fragment d'ADN inséré dans l'haplotype cible a été entreprise en retraçant sa présence dans la généalogie de notre cultivar d'étude R570. Les analyses sur une collection de 136 clones du genre *Saccharum* spp. suggèrent que cette insertion est ancienne et aurait été transmise aux cultivars modernes *via* des clones *Saccharum barberi*, des hybrides naturels entre *S. officinarum* et *S. spontaneum*.

Mots clés : polyploïdie, *Saccharum* spp., clonage positionnel, synténie, gène de résistance, canne à sucre.

### Structure of rust resistance locus in sugarcane, a highly polyploid species (2n=ca 12x=ca 115)

Modern sugarcane cultivars present one the most complex crop genome studied to date, mainly due to a very high degree of polyploidy, aneuploidy, together with an interspecific origin. They derive from the combination of two polyploid species: *Saccharum officinarum*, the domesticated sugar-producing species and *Saccharum spontaneum*, a vigorous wild species. The major gene *Bru1* conferring resistance to brown rust was identified in the modern cultivar R570. A map-based cloning approach had allowed developing a first physical map of the interest region, that encompasses seven hom(oe)ologous haplotypes which one corresponds to the target haplotype carrying *Bru1*. The target haplotype is made up of seven BAC clones that overlap only partially, including two gaps due to the presence of an inserted chromosome segment specific to this haplotype. In order to complete the physical map of the target haplotype, two strategies using BAC annotations were developed: (i) the first is based on the high level of gene conservation observed between hom(oe)ologous haplotypes and (ii) the second is based on the classical chromosome walking approach, by using markers flanking the two gaps. These strategies have enabled us to fill one of two gaps, partially cover the second gap and we suggest that the resistance gene would be present in the insertion. We have identified a candidate gene corresponding to a Serine/Threonine kinase located in the insertion. Further expression tests will be performed to validate this candidate gene. Meanwhile, the origin of the insertion located on the target haplotype was undertaken by tracing its origin in the R570 genealogy. Our analyses on 136 clones of a *Saccharum* spp. collection suggest that this insertion is old and has been transmitted to modern cultivars *via* *Saccharum barberi* clones, natural hybrids between *S. officinarum* and *S. spontaneum* species.

Keywords: polyploidy, *Saccharum* spp., map-based cloning, synteny, resistance gene, sugarcane.